1893

Multivariate Clustering Based on Entropy

U.S. GEOLOGICAL SURVEY BULLETIN 1893

I 19.3:

READEX FICHE best copy available

AVAILABILITY OF BOOKS AND MAPS OF THE U.S. GEOLOGICAL SURVEY

Instructions on ordering publications of the U.S. Geological Survey, along with prices of the last offerings, are given in the current-year issues of the monthly catalog "New Publications of the U.S. Geological Survey." Prices of available U.S. Geological Survey publications released prior to the current year are listed in the most recent annual "Price and Availability List." Publications that are listed in various U.S. Geological Survey catalogs (**see back inside cover**) but not listed in the most recent annual "Price and Availability List" are no longer available.

Prices of reports released to the open files are given in the listing "U.S. Geological Survey Open-File Reports," updated monthly, which is for sale in microfiche from the U.S. Geological Survey, Books and Open-File Reports Section, Federal Center, Box 25425, Denver, CO 80225. Reports released through the NTIS may be obtained by writing to the National Technical Information Service, U.S. Department of Commerce, Springfield, VA 22161; please include NTIS report number with inquiry.

Order U.S. Geological Survey publications **by mail** or **over the counter** from the offices given below.

BY MAIL

Books

Professional Papers, Bulletins, Water-Supply Papers, Techniques of Water-Resources Investigations, Circulars, publications of general interest (such as leaflets, pamphlets, booklets), single copies of Earthquakes & Volcanoes, Preliminary Determination of Epicenters, and some miscellaneous reports, including some of the foregoing series that have gone out of print at the Superintendent of Documents, are obtainable by mail from

U.S. Geological Survey, Books and Open-File Reports
Federal Center, Box 25425
Denver, CO 80225

Subscriptions to periodicals (Earthquakes & Volcanoes and Preliminary Determination of Epicenters) can be obtained ONLY from the

Superintendent of Documents
Government Printing Office
Washington, D.C. 20402

(Check or money order must be payable to Superintendent of Documents.)

Maps

For maps, address mail orders to

U.S. Geological Survey, Map Distribution
Federal Center, Box 25286
Denver, CO 80225

Residents of Alaska may order maps from

Alaska Distribution Section, U.S. Geological Survey,
New Federal Building - Box 12
101 Twelfth Ave., Fairbanks, AK 99701

OVER THE COUNTER

Books

Books of the U.S. Geological Survey are available over the counter at the following Geological Survey Public Inquiries Offices, all of which are authorized agents of the Superintendent of Documents:

- **WASHINGTON, D.C.**--Main Interior Bldg., 2600 corridor, 18th and C Sts., NW.
- **DENVER, Colorado**--Federal Bldg., Rm. 169, 1961 Stout St.
- **LOS ANGELES, California**--Federal Bldg., Rm. 7638, 300 N. Los Angeles St.
- **MENLO PARK, California**--Bldg. 3 (Stop 533), Rm. 3128, 345 Middlefield Rd.
- **RESTON, Virginia**--503 National Center, Rm. 1C402, 12201 Sunrise Valley Dr.
- **SALT LAKE CITY, Utah**--Federal Bldg., Rm. 8105, 125 South State St.
- **SAN FRANCISCO, California**--Customhouse, Rm. 504, 555 Battery St.
- **SPOKANE, Washington**--U.S. Courthouse, Rm. 678, West 920 Riverside Ave.
- **ANCHORAGE, Alaska**--Rm. 101, 4230 University Dr.
- **ANCHORAGE, Alaska**--Federal Bldg, Rm. E-146, 701 C St.

Maps

Maps may be purchased over the counter at the U.S. Geological Survey offices where books are sold (all addresses in above list) and at the following Geological Survey offices:

- **ROLLA, Missouri**--1400 Independence Rd.
- **DENVER, Colorado**--Map Distribution, Bldg. 810, Federal Center
- **FAIRBANKS, Alaska**--New Federal Bldg., 101 Twelfth Ave.

Multivariate Clustering Based on Entropy

By JOSEPH MOSES BOTBOL

Entropy-derived degrees of association proved as effective in describing degrees of association between geochemical variables as correlation coefficients of the raw data or their logarithms

U.S. GEOLOGICAL SURVEY BULLETIN 1893

DEPARTMENT OF THE INTERIOR
MANUEL LUJAN, JR., Secretary

U.S. GEOLOGICAL SURVEY
Dallas L. Peck, Director

Any use of trade, product, or firm names in this publication is for descriptive purposes only and does not imply endorsement by the U.S. Government

UNITED STATES GOVERNMENT PRINTING OFFICE: 1989

For sale by the
Books and Open-File Reports Section
U.S. Geological Survey
Federal Center, Box 25425
Denver, CO 80225

Library of Congress Cataloging in Publication Data

Botbol, Joseph Moses.
Multivariate clustering based on entropy / by Joseph Moses Botbol.
p. cm.—(U.S. Geological Survey bulletin ; 1893)
Supt. of Docs. no.: I 19.3:1893
1. Manganese nodules—North Pacific Ocean—Composition. 2. Cluster analysis. I. Title. II. Series.
QE390.2.M35B67 1990 89-600183
553.4'629—dc20 CIP

CONTENTS

Abstract 1
Introduction 1
    Acknowledgments 1
Relevant Theoretical Aspects of Entropy 1
Data Description 3
Entropy Computations 3
    Cobalt 3
    Nickel 3
    Manganese 3
    Iron 4
    Copper 4
    Zinc 4
    Lead 4
    Aluminum 4
    Silicon 4
    Depth 4
Cluster Analysis 4
    Correlation Coefficients of Raw Data 5
    Correlation Coefficients of Logs of Raw Data 5
    Uncertainty Coefficients of Elemental Enrichments 5
    Uncertainty Coefficients of Raw Data 5
Conclusions 5
References Cited 30

Figures

1. Scatter plots of cobalt versus nine variables 6
2. Perspective plots showing uncertainties of cobalt versus nine variables 7
3. Scatter plots of nickel versus nine variables 8
4. Perspective plots showing uncertainties of nickel versus nine variables 9
5. Scatter plots of manganese versus nine variables 10
6. Perspective plots showing uncertainties of manganese versus nine variables 11
7. Scatter plots of iron versus nine variables 12
8. Perspective plots showing uncertainties of iron versus nine variables 13
9. Scatter plots of copper versus nine variables 14
10. Perspective plots showing uncertainties of copper versus nine variables 15
11. Scatter plots of zinc versus nine variables 16
12. Perspective plots showing uncertainties of zinc versus nine variables 17
13. Scatter plots of lead versus nine variables 18
14. Perspective plots showing uncertainties of lead versus nine variables 19
15. Scatter plots of aluminum versus nine variables 20
16. Perspective plots showing uncertainties of aluminum versus nine variables 21
17. Scatter plots of silicon versus nine variables 22
18. Perspective plots showing uncertainties of silicon versus nine variables 23
19. Scatter plots of depth versus nine variables 24

20. Perspective plots showing uncertainties of depth versus nine variables **25**
21–24. Dendritic diagrams showing clusters of variables based on:
21. Linear correlation coefficients of raw data **26**
22. Linear correlation coefficients of logarithms of the raw data **27**
23. Uncertainty coefficients of elemental enrichments **28**
24. Uncertainty coefficients of the raw data **29**

Tables

1–10. Entropy computations for:
1. Cobalt versus nine variables **7**
2. Nickel versus nine variables **9**
3. Manganese versus nine variables **11**
4. Iron versus nine variables **13**
5. Copper versus nine variables **15**
6. Zinc versus nine variables **17**
7. Lead versus nine variables **19**
8. Aluminum versus nine variables **21**
9. Silicon versus nine variables **23**
10. Depth versus nine variables **25**
11. Correlation coefficients of raw data **26**
12. Correlation coefficients of logarithms of raw data **27**
13. Symmetrical uncertainties of elemental enrichments **28**
14. Symmetrical uncertainties of raw data **29**
15. Cluster analysis of linear correlation coefficients of raw data and their logarithmic transforms and uncertainty coefficients of elemental enrichments and raw data **30**

Multivariate Clustering Based on Entropy

By Joseph Moses Botbol

Abstract

To test the efficacy of entropy as a metric for degrees of association between pairs of variables, bivariate scatter plots of the sample depth and concentration of 9 elements within 1,200 "ferromanganese" nodules of the northern Pacific Ocean were classified into two-way contingency tables. Entropy attributes were computed for each table. The uncertainty coefficients (measures of the degree of association computed from other entropy attributes) of the raw data, data transformed to enrichments, linear correlation coefficients of the raw data, and raw data transformed to logarithms were input as similarity matrices to cluster analysis to determine the differences between the clusters of variables in each of the four input data groups.

Four similarity matrices were used as input to cluster analysis. Clusters of all input sources fell into three major groups of variables, with aluminum-silicon-depth, cobalt-lead, and nickel-copper-manganese-zinc common to their respective clusters in all four of the analyzed groups. The association of manganese-iron occurs only in the analysis of the raw data uncertainties.

To the extent of this study, entropy proved effective in describing degrees of association between variables in chemical analyses. Similar studies can be used as effective tools in the multivariate analysis of geochemical data.

INTRODUCTION

The various industrial, academic, and governmental agencies that are involved in marine geology or marine mineral resource studies are compiling and updating an ever-increasing number of marine geologic data bases. As the size and scope of these data bases increase, modern comprehensive geologic data analysis must embrace far more data and variables than ever before, thus resulting in a need to develop and use appropriate multivariate analytical methods to determine sensible and meaningful relationships between or among variables.

In the course of developing a hard-mineral resource model of the northern Pacific Ocean (Botbol and Evenden, 1988), chemical analyses of 9 elements for each of approximately 1,250 oxide-rich-phase marine nodules (also referred to as ferromanganese nodules) were evaluated statistically. Intermediate products of the analyses were scatter plots and two-way contingency tables of all pairs of variables. Linear correlation coefficients were initially computed for all pairs of variables, but these correlation coefficients were unacceptable as a final product because they were too easily corrupted by outliers. Ultimately, the relations between variables as portrayed in the scatter plots and contingency tables were visually classified as weakly, moderately, or strongly linear.

This report presents the results of a study undertaken to develop a technique for assessing the information content between pairs of variables so that their interrelations can be assessed for subsequent multivariate analysis. The primary inputs to this procedure are scatter plots of all pairs of variables. These are considered as order-disorder information diagrams, and entropy, as used in information theory, is the major metric used to describe relations between the variables in the diagrams.

Manuscript approved for publication April 20, 1989.

Acknowledgments

The author gratefully acknowledges R. Sinding-Larsen for suggestions regarding the use of entropy to generate similarity matrices, L.L. Montgomery for suggestions on various aspects of information theory, and G.I. Evenden for assistance with computing. A.L. Clark and J.R. Hein reviewed this manuscript.

RELEVANT THEORETICAL ASPECTS OF ENTROPY

Entropy is used to measure the degree of order or disorder of the components within a system. The systems may be as varied as the spatial distribution of electrons around the nucleus of an atom, the degree to which an evolving river system has come to equilibrium, or the amount of information and degree of association between variables in a two-way contingency table.

This study deals with the entropy characteristics of two-way contingency tables derived directly from bivariate

scatter plots of the chemical contents of marine ferromanganese nodules. A contingency table is composed of rows and columns whose intersections form cells that are analogous to the squares on a checkerboard. The row-column (X,Y) values of each datum determine the cell to which it belongs. The frequencies of points in the cells of their respective contingency tables constitute the basic data required to compute entropy. Equations 1 through 8 are from Press and others (1988).

The entropy of the whole of each contingency table is not used in computing the degrees of association of the variables, but it is a major entropy attribute useful in comparing scatter plots. It is calculated as follows:

$$H = -\sum_{i=1}^{I}\sum_{j=1}^{J} P_{i,j} \ln P_{i,j} \qquad (1)$$

where

H = total entropy of the contingency table,
$P_{i,j}$ = the proportion of total points in cell i,j,
I = number of columns, and
J = number of rows.

The logarithm is used to keep the proportional components of a given contingency table additive. The negation is used to keep the sum of the logarithms of fractions as a positive number. H is highest when all cells have an equal number of points, and H is lowest when all points are in one cell. The higher the disorder in the data, the higher the entropy, and conversely. This means that high degrees of association or dependence between variables are associated with low entropy values. [Singh (1966) refers to L. Brillouin who coined the term "negentropy" because of the inverse relation between entropy and information content.]

Four computation procedures are used to generate degrees of association between a pair of variables: (1) the individual entropy of each variable of the pair that composes the two-way contingency table (equations 2 and 3), (2) the entropy of each variable, taking into consideration the influence of the other (equations 4 and 5), (3) the coefficient of uncertainty for each variable, taking into consideration the influence of the other (equations 6 and 7), and (4) the weighted average of the coefficients of uncertainty (equation 8).

The individual entropies of X and Y in a contingency table of X versus Y are computed as follows:

$$H(Y) = -\sum_{j=1}^{J} P_j \ln P_j \qquad (2)$$

$$H(X) = -\sum_{i=1}^{I} P_i \ln P_i \qquad (3)$$

where

$H(Y)$ = entropy of Y,
$H(X)$ = entropy of X,
P_j = proportion of frequency of each cell relative to the frequency of its host row in the array,
P_i = proportion of frequency of each cell relative to the frequency of its host column in the array,
I = number of columns, and
J = number of rows.

Entropies for Y given X and for X given Y are computed as follows:

$$H(Y \mid X) = \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{P_{i,j} \ln P_{i,j}}{P_i} \qquad (4)$$

$$H(X \mid Y) = \sum_{j=1}^{J}\sum_{i=1}^{I} \frac{P_{i,j} \ln P_{i,j}}{P_j} \qquad (5)$$

where

$H(Y \mid X)$ = entropy of Y given X,
$H(X \mid Y)$ = entropy of X given Y,
P = proportion of points in cell,
I = number of columns, and
J = number of rows.

The computations that quantify the degrees of intervariable association are derived by computing the uncertainty coefficient, U, for each variable, taking into consideration the effect of the other variable of the pair. U indicates how much entropy is lost by one variable (that is, information that is gained) because of the presence of the other.

The formulas for $U(X \mid Y)$ and $U(Y \mid X)$ are as follows:

$$U(Y \mid X) = \frac{H(Y) - H(Y \mid X)}{H(Y)} \qquad (6)$$

$$U(X \mid Y) = \frac{H(X) - H(X \mid Y)}{H(X)} \qquad (7)$$

where

$U(Y \mid X)$ = uncertainty of Y given X,
$U(X \mid Y)$ = uncertainty of X given Y,
$H(Y)$ = entropy of Y,
$H(X)$ = entropy of X,
$H(Y \mid X)$ = entropy of Y given X, and
$H(X \mid Y)$ = entropy of X given Y.

U ranges from 0 to 1 where zero indicates no relation between X and Y, and 1 indicates complete predictability of one variable by the other.

To represent both $U(X \mid Y)$ and $U(Y \mid X)$ with a single number, the formula for symmetrical uncertainty, $U(X,Y)$, is computed as follows:

$$U(X,Y)=\frac{H(X)\ U(X \mid Y)+H(Y)\ U(Y \mid X)}{H(X)+H(Y)}. \qquad (8)$$

$U(X,Y)$ is the weighted average of both $U(X \mid Y)$ and $U(Y \mid X)$.

Entropy calculations are based on the information content of the data and do not preserve the sign of relations in the data. For example, two mirror image distributions having linear correlation coefficients of 0.7 and −0.7, respectively, will have the same entropy.

DATA DESCRIPTION

A northern Pacific Ocean oxide-rich-phase (ferromanganese) nodule subset of the Scripps Institute of Oceanography Sediment Data Bank (Frazer and Fisk, 1980) was used for this study. These data are those described by McKelvey and others (1983). The study area is bounded by 0° N., 40° N., 120° E., and 100° W. Each record in the data base is composed of multiple fields that describe one ocean bottom sample. The subset retrieved for this study is composed of the following variables: water depth above sample site, geographic coordinates, and concentration (in weight percent) of Co, Ni, Mn, Fe, Cu, Zn, Pb, Al, and Si. Not all variables are present in all samples.

Scatter plots were constructed for all pairs of variables, and each scatter plot was partitioned into a two-way contingency table. In each of the contingency tables, the number of classes for each variable was determined according to Sturges' Rule (from Huntsberger, 1963):

$$k= 1+3.3 \log n \qquad (9)$$

where

k = number of classes and
n = population size.

All class boundaries for the frequency distribution of each variable were arbitrarily set at equal intervals.

ENTROPY COMPUTATIONS

By using two-way contingency tables as input, entropy computations were made for all pairs of variables. The components and results of the entropy computations are presented herein for each variable and inclu le the original scatter plots of each variable versus each of the other variables, a tabular summary of all entropy computations, and a perspective plot of the uncertainty coefficients $U(X \mid Y)$, U(Y | X), and $U(X,Y)$.

The scatter plots present a synopsis of the order-disorder that exists in the raw data. Although the plots have been classified into two-way contingency tables for computation of entropy attributes, correlations can be visually estimated from the scatter plots for comparison with computed entropy values.

Tables 1 through 10 provide the complete range of entropy attributes computed for this study. Parenthetical table entries indicate the entropy calculation normalized by 1/log n. In the tables, "Percent data content" refers to the percentage of rows, columns, and cells for which there are data. For example, in table 1, cobalt versus nickel, every column and row of the contingency table contains at least one datum, whereas only 50 percent of the individual cells contain data. In all figures and tables, "de" stands for depth of water above sample site.

Perspective plots spatially portray the uncertainty coefficient of the subject variable as the dependent variable versus the other variables as independent variables ($U(X \mid Y)$), the subject variable as the independent variable versus the other variables as dependent variables ($U(Y \mid X)$), and the symmetrical uncertainty ($U(X,Y)$) of the subject variable versus the symmetrical uncertainties of the others. These diagrams provide a rapid review of the salient entropy relations of each subject variable with all of the other variables.

Cobalt

In figure 1, the scatter plots of cobalt versus lead and iron appear to be the strongest visually assessable linear correlations. The uncertainty plot in figure 2 demonstrates unequivocally that the association of cobalt and lead is the strongest of all the associations with cobalt. Although the average uncertainty coefficient of aluminum is numerically the next value lower than lead, it is clustered with the other variables. Zinc appears to be slightly removed from the major group of variables, but its unique position is in the low value side of the plot.

Nickel

Nickel scatter plots (fig. 3) show that nickel has a strong linear correlation with manganese, iron, copper, and zinc. These strong relations are also true of the uncertainties shown in figure 4, where copper has the highest degree of association with nickel. Associations with manganese, iron, and zinc are well defined but of lower value. Other variables are grouped near the origin of the figure (0,0) and show no significant degree of association.

Manganese

The three strong, visually assessable, linear correlations in figure 5 are those of manganese versus zinc, nickel,

and copper. These correlations are borne out by the uncertainty coefficients in figure 6, which set these variables apart from the other elements. Other variables are all grouped near the origin of the figure (0,0).

Iron

Figure 7 shows strong linear correlation of iron with nickel and copper and its moderate correlation with zinc and lead. These relations are clearly depicted in figure 8. The degrees of association implied by the uncertainties of the other variables are distinctly lower.

Copper

Copper has a very strong linear relation to nickel and moderate to weak correlations with manganese, iron, and zinc (fig. 9). Figure 10 shows an exceptionally high degree of association between nickel and copper. Although lower than the copper-nickel uncertainty coefficient, those of copper to iron, manganese, and zinc are also high.

Zinc

Figure 11 shows moderate to strong linear correlation between zinc and nickel, manganese, and copper. These variables also have the highest degrees of association with zinc (fig. 12). In figure 12, the almost horizontal alignment of the variables is due to the overall high entropy in the Y direction, which is seen in figure 11 where a narrow range of zinc concentrations occurs with a very broad range of data values of the other elements.

Lead

Figure 13 shows a weak to moderate correlation between lead and cobalt, iron, copper, and zinc. Uncertainty coefficients in figure 14 show that lead has a high degree of association with zinc, iron, copper, and cobalt. Other variables show no significant relations.

Aluminum

Aluminum's only strong correlation is with silicon (fig. 15). This correlation is well emphasized in the plot of the pair's high uncertainty coefficient (fig. 16). At the scale of figure 16, the high value of the uncertainty coefficient of the aluminum and silicon pair seriously diminishes the graphic portrayal of any other significant relations existing between the other variables and aluminum.

Silicon

Silicon is very significantly associated with aluminum in both the scatter plots of figure 17 and the plot of the uncertainty coefficients in figure 18. As with aluminum, the relation between silicon and aluminum dominates any relations between silicon and the other variables.

Depth

In figure 19, depth shows very weak correlation with copper and almost no correlation with any of the other variables. Also, there appears to be a vestige of a second population in each of the scatter plots. The uncertainties in figure 20 show three distinct clusters of variables associated with depth: (1) copper, (2) iron, aluminum, nickel, lead, and silicon, and (3) zinc, cobalt, and manganese.

CLUSTER ANALYSIS

Up to this point in the analysis, it has been shown that uncertainty coefficients of the variables more or less agree with a visual estimate of their linear correlations and, when displayed as perspective pole diagrams, show a striking visual representation of information content. The next logical step in the comparison of entropy-derived similarities with those derived by traditional methods is to cluster the variables.

Cluster analysis is a widely accepted numerical taxonomy method of partitioning the variables that compose an input similarity matrix to classify the variables into groups of maximum similarity. If the groups of variables determined from the entropy-based similarity matrices sensibly simulate those of accepted similarity matrices, such as matrices of correlation coefficients, similarity between the input matrices is implicit, and further methodological development and utilization are justified.

Four similarity matrices were used as input to cluster analysis: (1) product-moment linear correlation coefficients between all pairs of variables in the raw data; (2) product-moment linear correlation coefficients between all pairs of variables in the raw data as transformed to their logarithms; (3) uncertainty coefficients of all pairs of variables in the raw data as transformed to their elemental enrichments (Botbol, 1987); and (4) uncertainty coefficients of all pairs of variables in the raw data. These four matrices provide a basis for comparison between the entropy-derived similarities and those derived by traditional methods. The results of this analysis are presented in four dendritic diagrams, figures 21 through 24, derived from the source data presented in tables 11 through 14, respectively. Final summary of the data clustering is presented in table 15, where the component variables of each cluster for each input similarity matrix are grouped.

Included in the cluster analysis is a statement of the cophenetic correlation. This is a measure of the degree of distortion introduced into the dendritic diagram. If the value is less than 0.8, the cluster analysis is subject to question (Davis, 1973, p. 466). All cluster analyses were performed by using the algorithms presented in the CLUSTR program (Davis, 1973, p. 467).

Correlation coefficients of raw data

Table 11 is the matrix of product-moment linear correlation coefficients of the raw data used in the cluster analysis procedure. Intuitive partitioning of the dendritic diagram in figure 21 shows three major clusters of variables: (1) cobalt, lead, and iron; (2) nickel, copper, manganese, and zinc; and (3) aluminum, silicon, and depth.

Correlation coefficients of logs of raw data

Table 12, the second cluster analysis similarity matrix, is composed of the product-moment linear correlation coefficients of the logarithms of the raw data. Figure 22 shows three major clusters of variables that are the same as those in the first analysis. Interestingly, the ordinal aspects of the clusters and the cluster component variables for the first two analyses are identical.

Uncertainty coefficients of elemental enrichments

The third cluster analysis (fig. 23) is based on the similarity matrix of uncertainty coefficients of elemental enrichments shown in table 13. The same three major clusters of variables occur as in the analysis of the raw data. Note that, although the order of some of the variables has changed, the variables within the clusters are the same.

Uncertainty coefficients of raw data

The fourth and final cluster analysis uses as input the uncertainty coefficient matrix for the raw data shown in table 14. Three distinct clusters result: (1) cobalt and lead; (2) nickel, copper, iron, manganese, and zinc; and (3) aluminum, silicon, and depth (fig. 24).

CONCLUSIONS

Table 15 summarizes the four data groups that were subjected to cluster analysis. There are three distinct clusters in each of the analyses. In the table, variables for each group are arranged so that the leftmost variable has the highest value.

The most revealing aspect of table 15 is the shifting of iron from cluster I of the correlation coefficients of the raw data, their logarithmic transforms, and elemental enrichment uncertainty coefficients partitions to cluster II of the raw data uncertainty coefficients. This is the only partition in which iron is grouped with manganese. Interestingly, the most fundamental attribute of the data used in this study is the association of the two dominant elements that compose ferromanganese nodules. The uncertainty coefficients of the raw data compose the only input similarity matrix that reflects this association.

The strongest and most pervasive cluster in the entire analysis is cluster III, which groups aluminum, silicon, and depth This composition undoubtedly reflects the geologic environment of the deep sea nodules in terms of nodule core material and in terms of the geologic distribution of the aluminosilicate fraction in the north Pacific. All three components are equivalent at an ordinal level and show exceptionally strong degrees of association in the plots of uncertainty coefficients.

Two sets of variables that occurred within the same clusters during all of the analyses are cobalt and lead in cluster I and nickel, copper, manganese, and zinc in cluster II. All show strong degrees of association in the plots of the uncertainty coefficients.

Another significant aspect of the results of this study is that the uncertainty coefficients of the raw data ordinally correspond almost exactly with the visual estimation of their linear correlation coefficients. In and of itself, this correspondence is not categorical proof of the efficacy of uncertainty coefficients as replacements for linear correlation coefficients, but it does establish significant credibility in entropy attributes as viable metrics of degrees of association between variables.

With respect to establishing credibility of multivariate methods that use entropy-based attributes of data, this study has shown that entropy attributes of the raw data compare favorably with data attributes that are determined by linear correlation of both raw data and their log transforms and with the entropy-derived attributes of enrichment transforms.

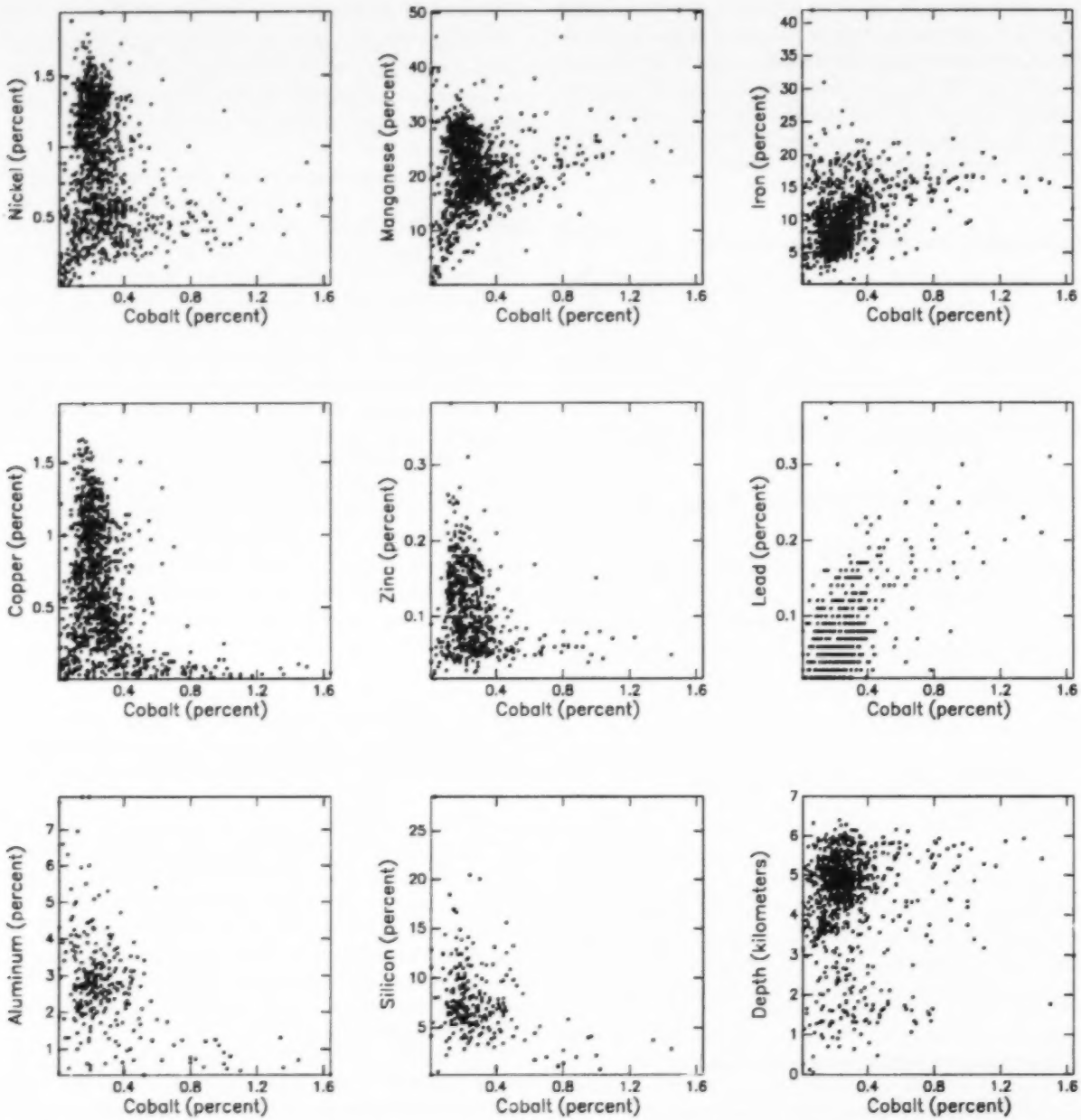

Figure 1. Scatter plots of Co content versus content of Ni, Mn, Fe, Cu, Zn, Pb, Al, and Si and depth of water above the sample site.

Table 1. Entropy computations for cobalt versus nine variables

[Parenthetical entries indicate entropy calculation normalized by 1/log *n*. de, depth of water above the sample site]

	Ni	Mn	Fe	Cu	Zn	Pb	Al	Si	de
Entropy of *X*	1.32 (.55)	1.32 (.55)	1.32 (.55)	1.31 (.54)	1.21 (.50)	1.33 (.55)	1.47 (.61)	1.38 (.57)	1.35 (.56)
Entropy of *Y*	2.18 (.91)	1.74 (.72)	1.49 (.62)	2.15 (.89)	0.76 (.33)	1.45 (.63)	1.66 (.75)	1.44 (.65)	1.93 (.80)
Entropy of (*X*,*Y*)	3.36 (.70)	2.93 (.61)	2.66 (.55)	3.28 (.68)	1.92 (.40)	2.54 (.54)	2.94 (.64)	2.69 (.58)	3.16 (.66)
Entropy of (*Y* | *X*)	2.03 (.84)	1.61 (.67)	1.34 (.55)	1.97 (.82)	0.70 (.30)	1.20 (.52)	1.47 (.67)	1.31 (.59)	1.81 (.75)
Entropy of (*X* | *Y*)	1.17 (.49)	1.19 (.49)	1.17 (.48)	1.13 (.47)	1.15 (.48)	1.09 (.45)	1.28 (.53)	1.24 (.52)	1.23 (.51)
Uncertainty of (*Y* | *X*)	0.06	0.07	0.10	0.08	0.08	0.16	0.11	0.09	0.06
Uncertainty of (*X* | *Y*)	.11	.09	.11	.13	.05	.18	.12	.09	.08
Uncertainty of (*X*,*Y*)	.08	.08	.10	.10	.06	.17	.11	.09	.07
No. of points	1,164	1,153	1,146	1,158	688	624	301	258	1,058
Percent data content *									
Columns	100	90	100	100	90	100	90	81	100
Rows	100	90	72	90	50	100	100	77	100
Cells	50	40	38	39	18	42	37	30	55

* Percentage of rows, columns, and cells for which there are data.

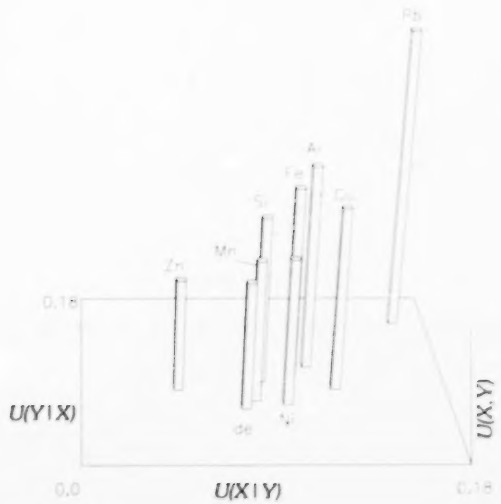

Figure 2. Perspective plot of uncertainties of Co versus uncertainties of Ni, Mn, Fe, Cu, Zn, Pb, Al, Si, and depth of water above the sample site (de). $U(Y \mid X)$, $U(X \mid Y)$, and $U(X,Y)$ are uncertainty coefficients.

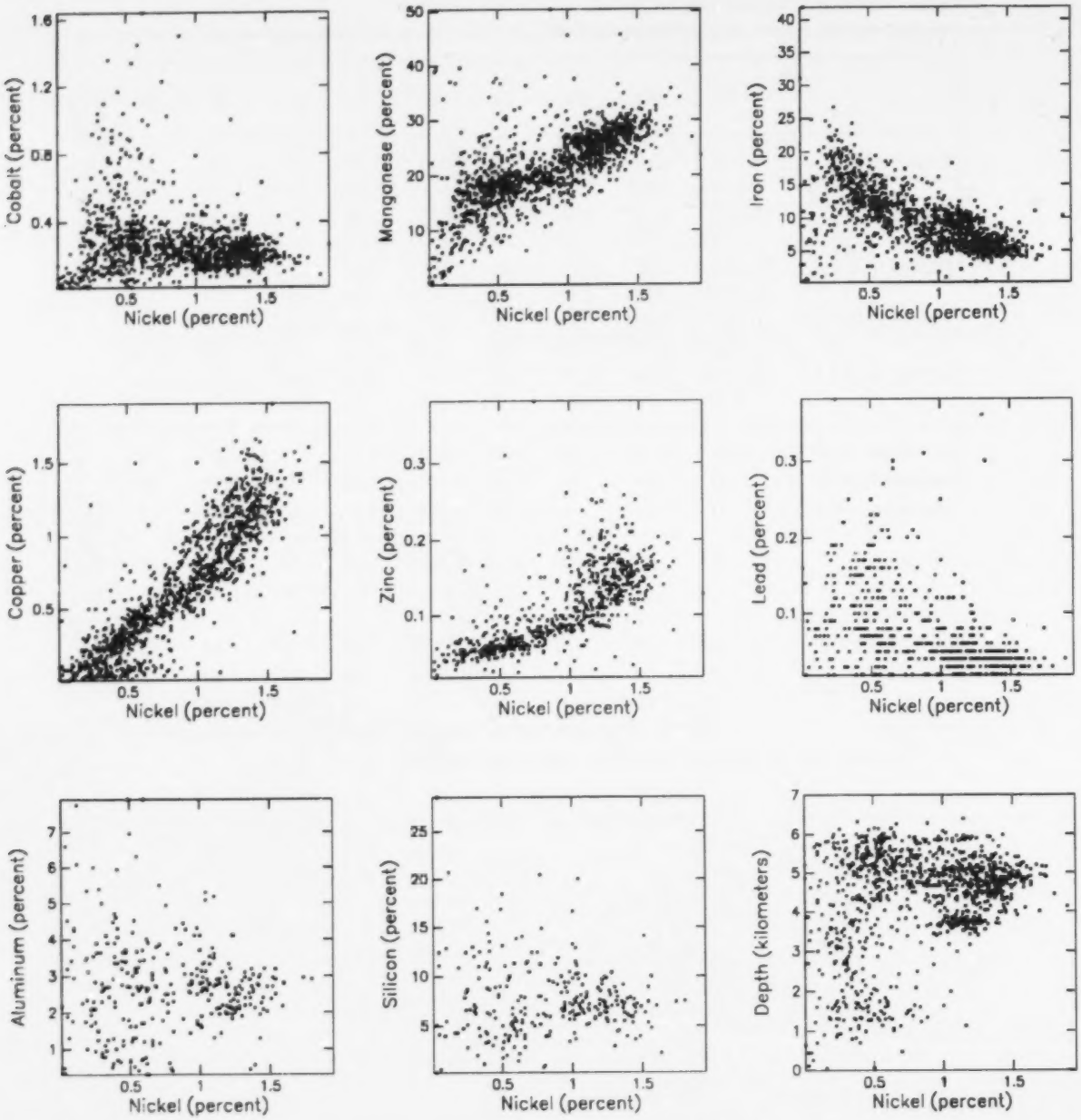

Figure 3. Scatter plots of Ni content versus content of Co, Mn, Fe, Cu, Zn, Pb, Al, and Si and depth of water above the sample site.

Table 2. Entropy computations for nickel versus nine variables

[Parenthetical entries indicate entropy calculation normalized by 1/log *n*. de, depth of water above the sample site]

	Co	Mn	Fe	Cu	Zn	Pb	Al	Si	de
Entropy of X	2.18 (.91)	2.18 (.90)	2.17 (.90)	2.18 (.91)	2.13 (.89)	2.19 (.91)	2.17 (.90)	2.18 (.91)	2.18 (.91)
Entropy of Y	1.32 (.55)	1.73 (.72)	1.49 (.62)	2.15 (.89)	0.76 (.33)	1.44 (.62)	1.67 (.76)	1.45 (.66)	1.96 (.81)
Entropy of (X,Y)	3.36 (.70)	3.39 (.70)	3.23 (.67)	3.54 (.73)	2.57 (.54)	3.41 (.72)	3.67 (.80)	3.46 (.75)	3.89 (.81)
Entropy of $(Y \mid X)$	1.17 (.49)	1.21 (.50)	1.05 (.44)	1.35 (.56)	0.43 (.19)	1.22 (.53)	1.49 (.68)	1.27 (.58)	1.71 (.71)
Entropy of $(X \mid Y)$	2.03 (.84)	1.65 (.69)	1.73 (.72)	1.38 (.57)	1.80 (.75)	1.96 (.82)	2.00 (.83)	2.00 (.83)	1.93 (.80)
Uncertainty of $(Y \mid X)$	0.11	0.30	0.29	0.37	0.42	0.15	0.10	0.12	0.12
Uncertainty of $(X \mid Y)$	.06	.24	.20	.36	.15	.10	.07	.08	.11
Uncertainty of (X,Y)	.08	.26	.23	.36	.22	.12	.09	.09	.12
No. of points	1,164	1,215	1,208	1,158	693	628	310	266	1,118
Percent data content *									
Columns	100	100	100	100	90	100	100	100	100
Rows	100	90	72	90	50	100	100	77	100
Cells	50	50	45	55	23	49	56	48	67

* Percentage of rows, columns, and cells for which there are data.

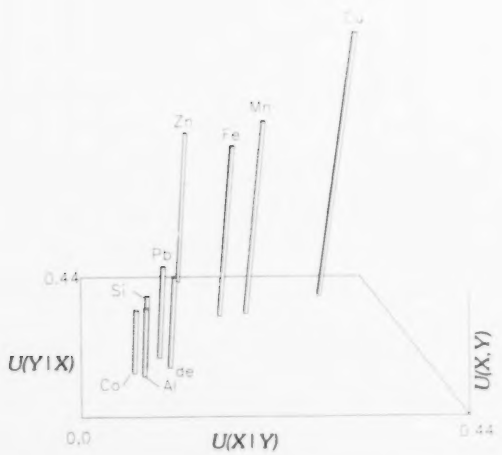

Figure 4. Perspective plot of uncertainties of Ni versus uncertainties of Co, Mn, Fe, Cu, Zn, Pb, Al, Si, and depth of water above the sample site (de). $U(Y \mid X)$, $U(X \mid Y)$, and $U(X,Y)$ are uncertainty coefficients.

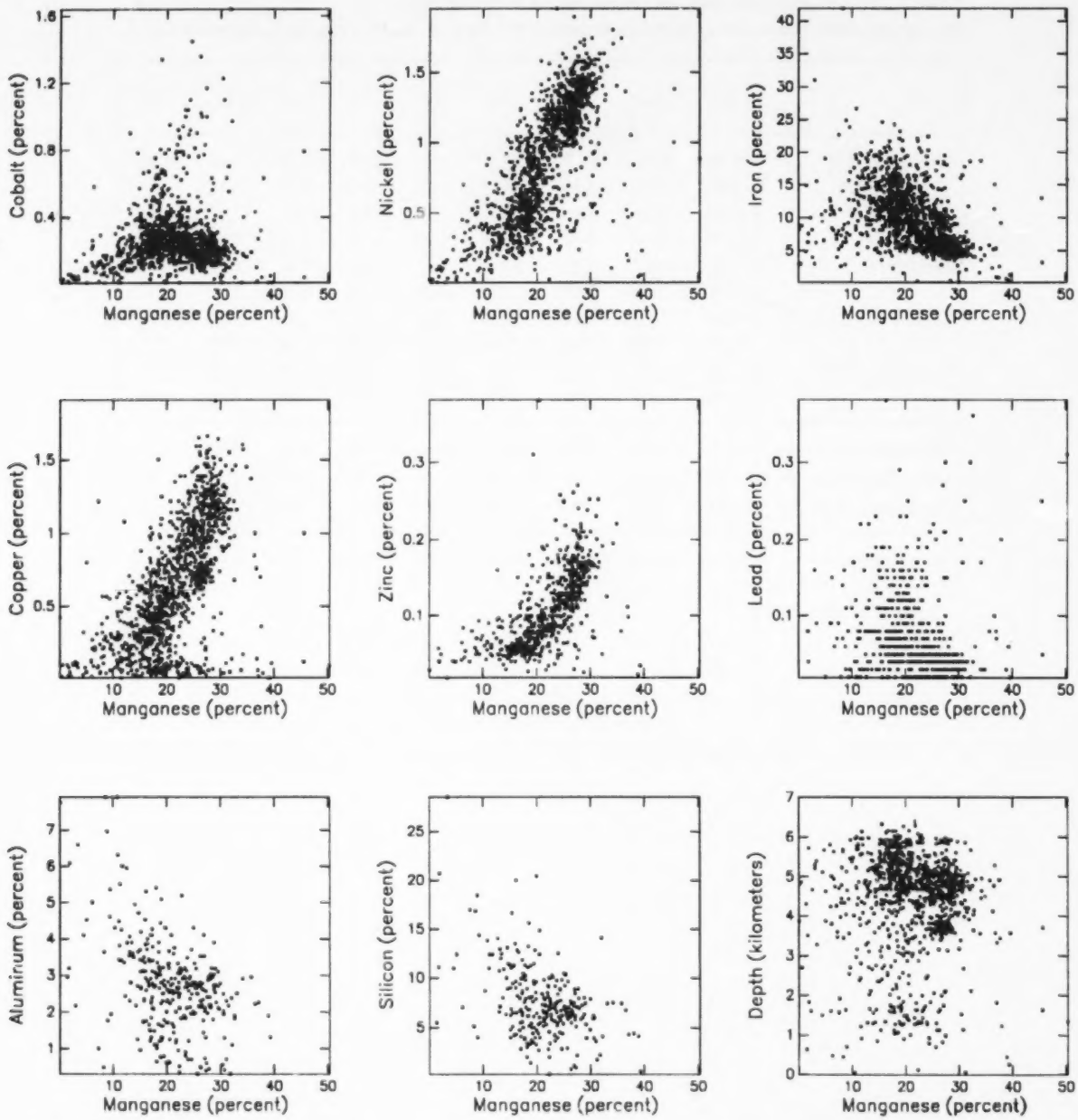

Figure 5. Scatter plots of Mn content versus content of Co, Ni, Fe, Cu, Zn, Pb, Al, and Si and depth of water above the sample site.

Table 3. Entropy computations for manganese versus nine variables

[Parenthetical entries indicate entropy calculation normalized by 1/log *n*. de, depth of water above the sample site]

	Co	Ni	Fe	Cu	Zn	Pb	Al	Si	de
Entropy of X	1.74	1.73	1.73	1.73	1.63	1.75	1.78	1.71	1.76
	(.72)	(.72)	(.72)	(.72)	(.68)	(.73)	(.74)	(.71)	(.73)
Entropy of Y	1.32	2.18	1.49	2.15	0.76	1.44	1.67	1.45	1.96
	(.55)	(.90)	(.62)	(.89)	(.33)	(.62)	(.76)	(.66)	(.81)
Entropy of (X,Y)	2.93	3.39	3.01	3.50	2.12	3.04	3.27	2.98	3.59
	(.61)	(.70)	(.62)	(.73)	(.45)	(.64)	(.71)	(.64)	(.74)
Entropy of $(Y \mid X)$	1.19	1.65	1.27	1.76	0.48	1.29	1.49	1.26	1.83
	(.49)	(.69)	(.53)	(.73)	(.21)	(.56)	(.68)	(.57)	(.76)
Entropy of $(X \mid Y)$	1.61	1.21	1.51	1.34	1.35	1.60	1.60	1.52	1.63
	(.67)	(.50)	(.63)	(.56)	(.56)	(.66)	(.67)	(.63)	(.68)
Uncertainty $(Y \mid X)$	0.09	0.24	0.14	0.18	0.36	0.10	0.10	0.13	0.06
Uncertainty $(X \mid Y)$	.07	.30	.12	.22	.17	.08	.09	.11	.07
Uncertainty (X,Y)	.08	.26	.13	.19	.23	.09	.10	.12	.06
No. of points	1,153	1,215	1,208	1,205	693	627	311	267	1,108
Percent data content *									
Columns	90	90	90	90	81	90	81	81	90
Rows	90	100	72	90	50	100	100	77	100
Cells	40	50	44	52	18	49	48	35	69

* Percentage of rows, columns, and cells for which there are data.

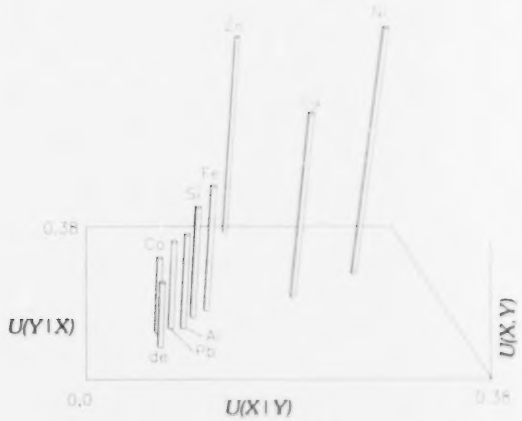

Figure 6. Perspective plot of uncertainties of Mn versus uncertainties of Co, Ni, Fe, Cu, Zn, Pb, Al, Si, and depth of water above the sample site (de). $U(Y \mid X)$, $U(X \mid Y)$, and $U(X,Y)$ are uncertainty coefficients.

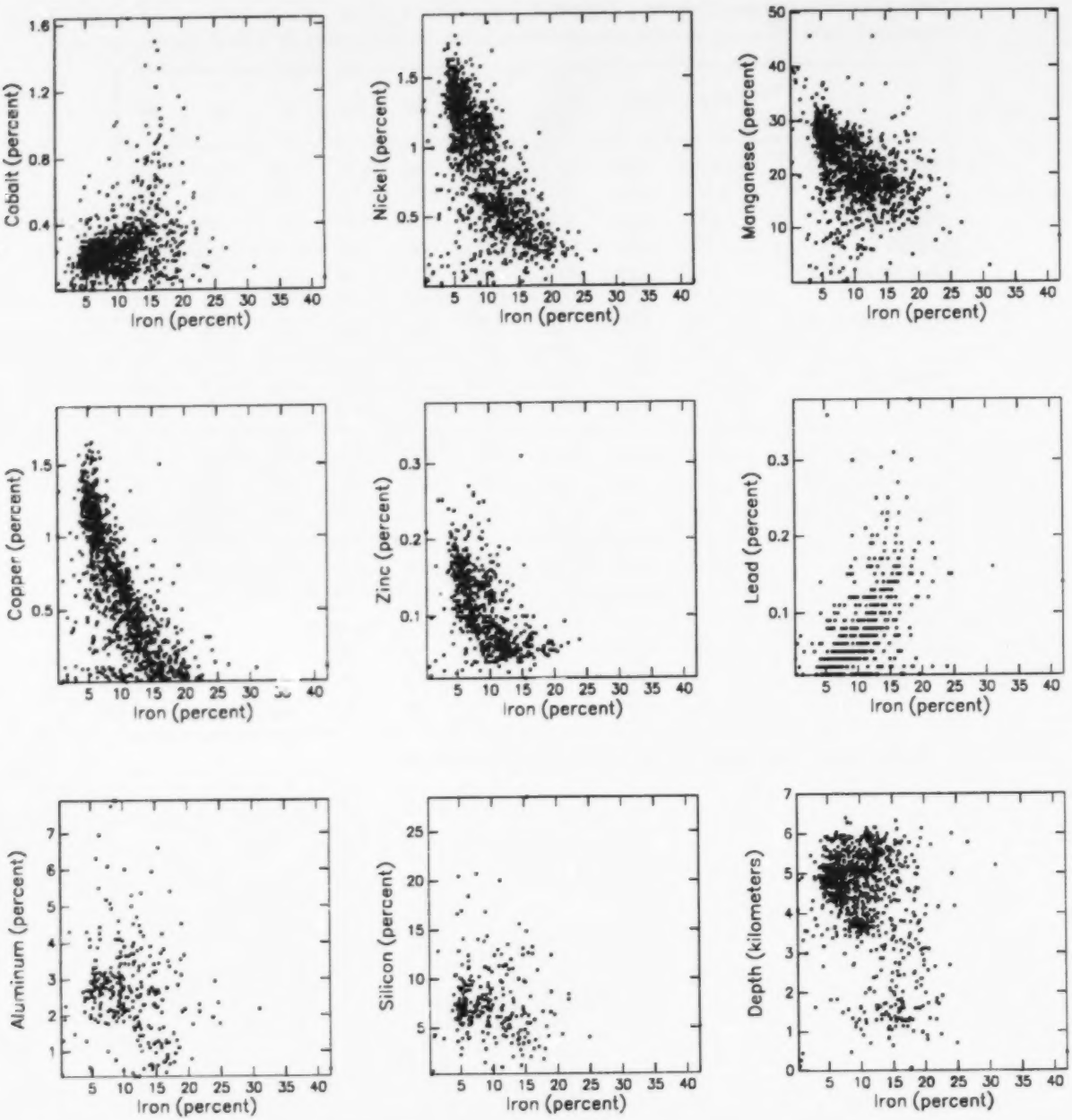

Figure 7. Scatter plots of Fe content versus content of Co, Ni, Mn, Cu, Zn, Pb, Al, and Si and depth of water above the sample site.

Table 4. Entropy computations for iron versus nine variables

[Parenthetical entries indicate entropy calculation normalized by 1/log n. de, depth of water above the sample site]

	Co	Ni	Mn	Cu	Zn	Pb	Al	Si	de
Entropy of X	1.49 (.62)	1.49 (.62)	1.49 (.62)	1.49 (.62)	1.38 (.57)	1.48 (.62)	1.57 (.65)	1.44 (.60)	1.50 (.62)
Entropy of Y	1.32 (.55)	2.17 (.90)	1.73 (.72)	2.15 (.89)	0.767 (.33)	1.44 (.62)	1.67 (.76)	1.45 (.66)	1.96 (.81)
Entropy of (X,Y)	2.66 (.55)	3.23 (.67)	3.01 (.62)	3.16 (.66)	1.98 (.42)	2.66 (.56)	3.10 (.67)	2.79 (.60)	3.24 (.67)
Entropy of $(Y \mid X)$	1.17 (.48)	1.73 (.72)	1.51 (.63)	1.67 (.70)	0.609 (.26)	1.17 (.51)	1.53 (.69)	1.34 (.61)	1.73 (.72)
Entropy of $(X \mid Y)$	1.34 (.55)	1.05 (.44)	1.27 (.53)	1.01 (.42)	1.22 (.50)	1.22 (.50)	1.43 (.59)	1.33 (.55)	1.27 (.53)
Uncertainty $(Y \mid X)$	0.11	0.201	0.12	0.22	0.21	0.18	0.08	0.07	0.11
Uncertainty $(X \mid Y)$	.10	.295	.14	.31	.12	.18	.08	.07	.15
Uncertainty (X,Y)	.10	.238	.13	.264	.15	.18	.08	.07	.13
No. of points	1,146	1,208	1,208	1,198	694	626	310	266	1,100
Percent data content *									
columns	72	72	72	72	63	72	72	63	72
rows	100	100	90	90	50	100	100	77	100
cells	38	45	44	38	15	34	42	31	53

* Percentage of rows, columns, and cells for which there are data.

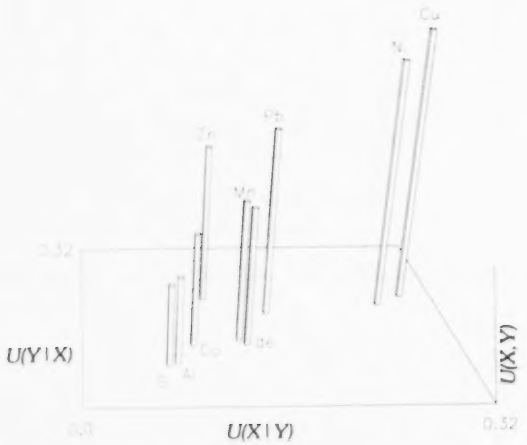

Figure 8. Perspective plot of uncertainties of Fe versus uncertainties of Co, Ni, Mn, Cu, Zn, Pb, Al, Si, and depth of water above the sample site (de). $U(Y \mid X)$, $U(X \mid Y)$, and $U(X,Y)$ are uncertainty coefficients.

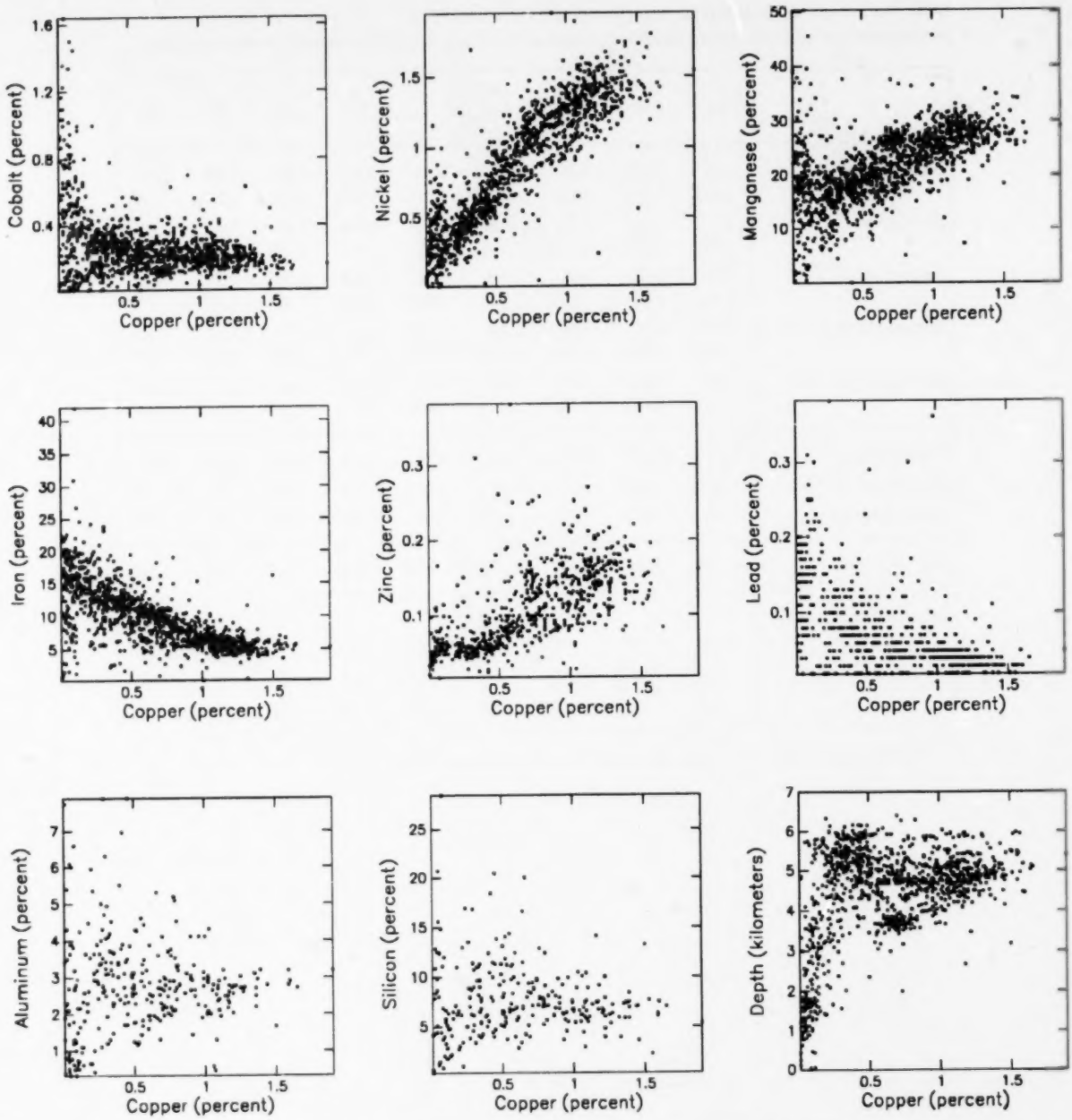

Figure 9. Scatter plots of Cu content versus content of Co, Ni, Mn, Fe, Zn, Pb, Al, and Si and depth of water above the sample site.

Table 5. Entropy computations for copper versus nine variables

[Parenthetical entries indicate entropy calculation normalized by 1/log *n*. de, depth of water above the sample site]

	Co	Ni	Mn	Fe	Zn	Pb	Al	Si	de
Entropy of X	2.15 (.89)	2.15 (.89)	2.15 (.89)	2.15 (.89)	2.17 (.90)	2.16 (.90)	2.05 (.85)	2.15 (.89)	2.15 (.89)
Entropy of Y	1.31 (.54)	2.18 (.91)	1.73 (.72)	1.49 (.62)	0.7691 (.33)	1.45 (.62)	1.66 (.75)	1.45 (.66)	1.95 (.81)
Entropy of (X,Y)	3.28 (.68)	3.54 (.73)	3.50 (.73)	3.16 (.66)	2.68 (.57)	3.31 (.70)	3.47 (.75)	3.45 (.75)	3.72 (.77)
Entropy of $(Y \mid X)$	1.13 (.47)	1.38 (.57)	1.34 (.56)	1.01 (.42)	0.51 (.22)	1.15 (.50)	1.41 (.64)	1.29 (.59)	1.56 (.65)
Entropy of $(X \mid Y)$	1.97 (.82)	1.35 (.56)	1.76 (.73)	1.67 (.70)	1.91 (.79)	1.86 (.77)	1.80 (.75)	1.99 (.83)	1.76 (.73)
Uncertainty $(Y \mid X)$	0.13	0.36	0.22	0.31	0.33	0.20	0.15	0.11	0.19
Uncertainty $(X \mid Y)$	.08	.37	.18	.22	.11	.13	.12	.07	.17
Uncertainty (X,Y)	.10	.36	.19	.26	.17	.16	.13	.08	.18
No. of points	1,158	1,158	1,205	1,198	692	625	305	261	1,107
Percent data content *									
columns	90	90	90	90	90	90	90	90	90
rows	100	100	90	72	50	100	100	77	100
cells	39	55	52	38	25	42	47	46	52

* Percentage of rows, columns, and cells for which there are data.

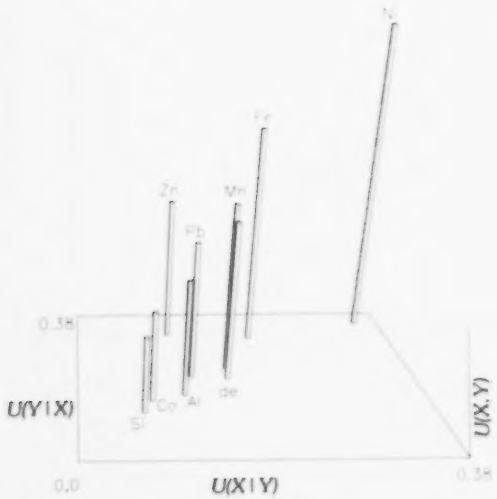

Figure 10. Perspective plot of uncertainties of Cu versus uncertainties of Co, Ni, Mn, Fe, Zn, Pb, Al, Si, and depth of water above the sample site (de). $U(Y \mid X)$, $U(X \mid Y)$, and $U(X,Y)$ are uncertainty coefficients.

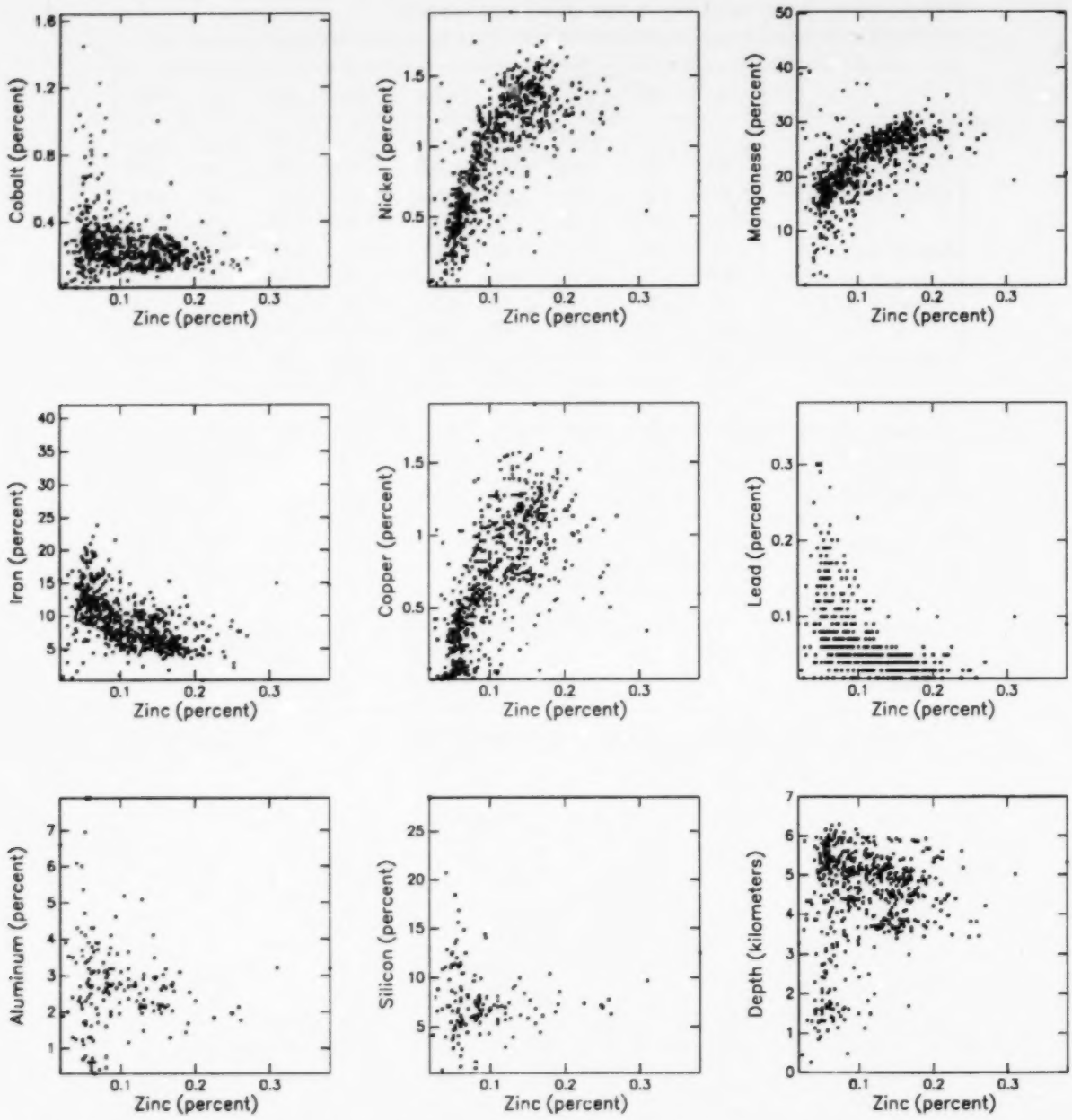

Figure 11. Scatter plots of Zn content versus content of Co, Ni, Mn, Fe, Cu, Pb, Al, and Si and depth of water above the sample site.

Table 6. Entropy computations for zinc versus nine variables

[Parenthetical entries indicate entropy calculation normalized by 1/log *n*. de, depth of water above the sample site]

	Co	Ni	Mn	Fe	Cu	Pb	Al	Si	de
Entropy of *X*	0.76 (.33)	0.76 (.33)	0.76 (.33)	0.76 (.33)	0.76 (.33)	0.76 (.33)	0.77 (.33)	0.63 (.27)	0.76 (.33)
Entropy of *Y*	1.21 (.50)	2.13 (.89)	1.63 (.68)	1.38 (.57)	2.17 (.90)	1.38 (.60)	1.59 (.72)	1.43 (.65)	1.96 (.81)
Entropy of (*X*,*Y*)	1.9 (.40)	2.57 (.54)	2.12 (.45)	1.98 (.42)	2.68 (.57)	1.96 (.42)	2.26 (.50)	2.02 (.45)	2.60 (.55)
Entropy of (*Y* | *X*)	1.15 (.48)	1.80 (.75)	1.35 (.56)	1.22 (.50)	1.91 (.79)	1.20 (.52)	1.49 (.67)	1.39 (.63)	1.84 (.76)
Entropy of (*X* | *Y*)	0.70 (.30)	0.43 (.19)	0.48 (.21)	0.60 (.26)	0.51 (.22)	0.57 (.24)	0.67 (.29)	0.58 (.25)	0.64 (.27)
Uncertainty (*Y* | *X*)	0.05	0.15	0.17	0.12	0.11	0.13	0.06	0.03	0.06
Uncertainty (*X* | *Y*)	.08	.42	.36	.21	.33	.24	.13	.06	.15
Uncertainty (*X*,*Y*)	.06	.22	.23	.15	.17	.17	.08	.04	.08
No. of points	688	693	693	694	692	532	178	144	646
Percent data content *									
columns	50	50	50	50	50	50	50	40	50
rows	90	90	81	63	90	80	88	77	100
cells	18	23	18	15	25	16	20	16	23

* Percentage of rows, columns, and cells for which there are data.

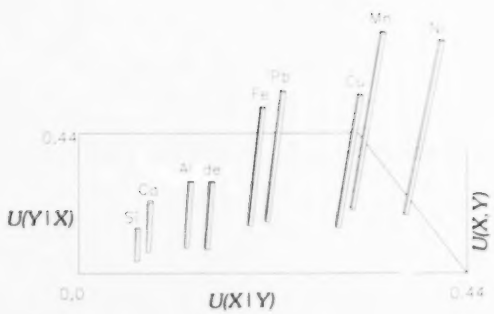

Figure 12. Perspective plot of uncertainties of Zn versus uncertainties of Co, Ni, Mn, Fe, Cu, Pb, Al, Si, and depth of water above the sample site (de). $U(Y \mid X)$, $U(X \mid Y)$, and $U(X,Y)$ are uncertainty coefficients.

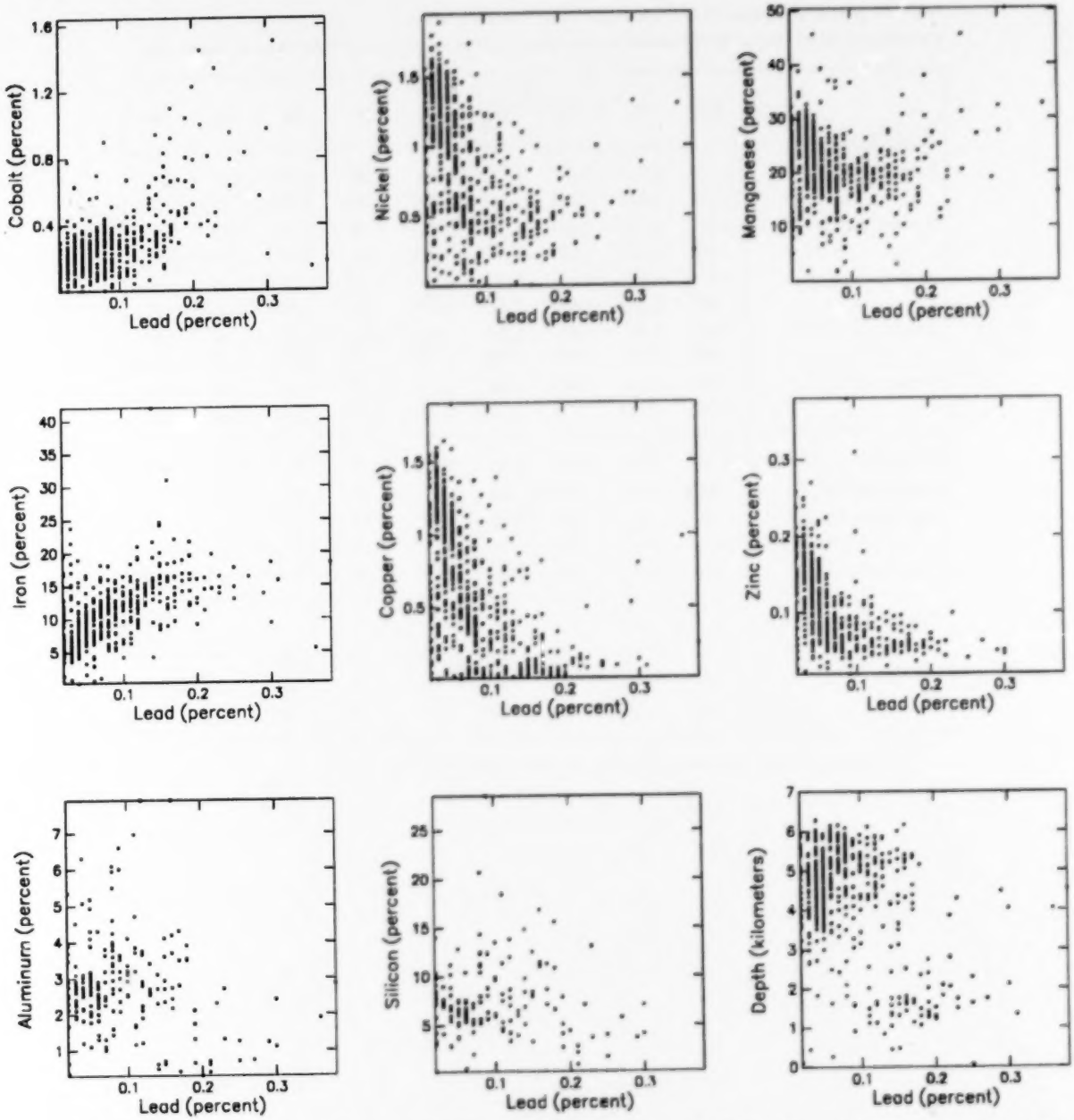

Figure 13. Scatter plots of Pb content versus content of Co, Ni, Mn, Fe, Cu, Zn, Al, and Si and depth of water above the sample site.

Table 7. Entropy computations for lead versus nine variables

[Parenthetical entries indicate entropy calculation normalized by 1/log *n*. de, depth of water above the sample site]

	Co	Ni	Mn	Fe	Cu	Zn	Al	Si	de
Entropy of *X*	1.45 (.63)	1.44 (.62)	1.44 (.62)	1.44 (.62)	1.45 (.62)	1.38 (.60)	1.67 (.72)	1.71 (.74)	1.46 (.63)
Entropy of *Y*	1.33 (.55)	2.19 (.91)	1.75 (.73)	1.48 (.62)	2.16 (.90)	0.76 (.33)	1.69 (.76)	1.51 (.68)	1.98 (.82)
Entropy of (*X*,*Y*)	2.54 (.54)	3.41 (.72)	3.04 (.64)	2.66 (.56)	3.31 (.70)	1.96 (.42)	3.17 (.70)	3.07 (.68)	3.22 (.68)
Entropy of (*Y* | *X*)	1.09 (.45)	1.96 (.82)	1.60 (.66)	1.22 (.50)	1.86 (.77)	0.57 (.24)	1.50 (.68)	1.35 (.61)	1.76 (.73)
Entropy of (*X* | *Y*)	1.20 (.52)	1.22 (.53)	1.29 (.56)	1.17 (.51)	1.15 (.50)	1.20 (.52)	1.48 (.64)	1.56 (.67)	1.24 (.54)
Uncertainty (*Y* | *X*)	0.18	0.10	0.08	0.18	0.13	0.24	0.11	0.10	0.10
Uncertainty (*X* | *Y*)	.16	.15	.10	.18	.20	.13	.11	.09	.14
Uncertainty (*X*,*Y*)	.17	.12	.09	.18	.16	.17	.11	.09	.12
No. of points	624	628	627	626	625	532	203	156	616
Percent data content *									
columns	100	100	100	100	100	80	100	90	100
rows	100	100	90	72	90	50	88	77	100
cells	42	49	49	34	42	16	42	41	58

* Percentage of rows, columns, and cells for which there are data.

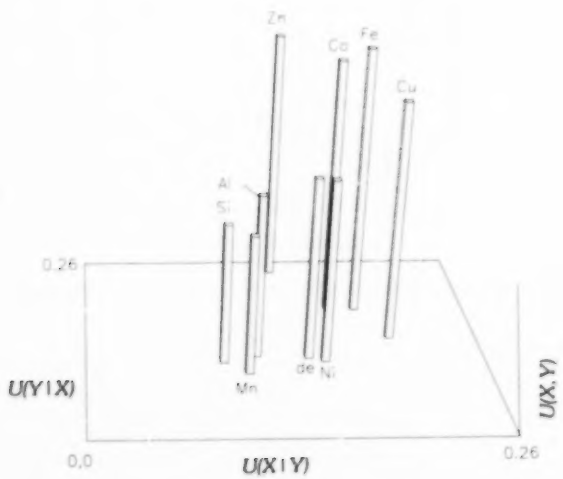

Figure 14. Perspective plot of uncertainties of Pb versus uncertainties of Co, Ni, Mn, Fe, Cu, Zn, Al, Si, and depth of water above the sample site (de). $U(Y \mid X)$, $U(X \mid Y)$, and $U(X,Y)$ are uncertainty coefficients.

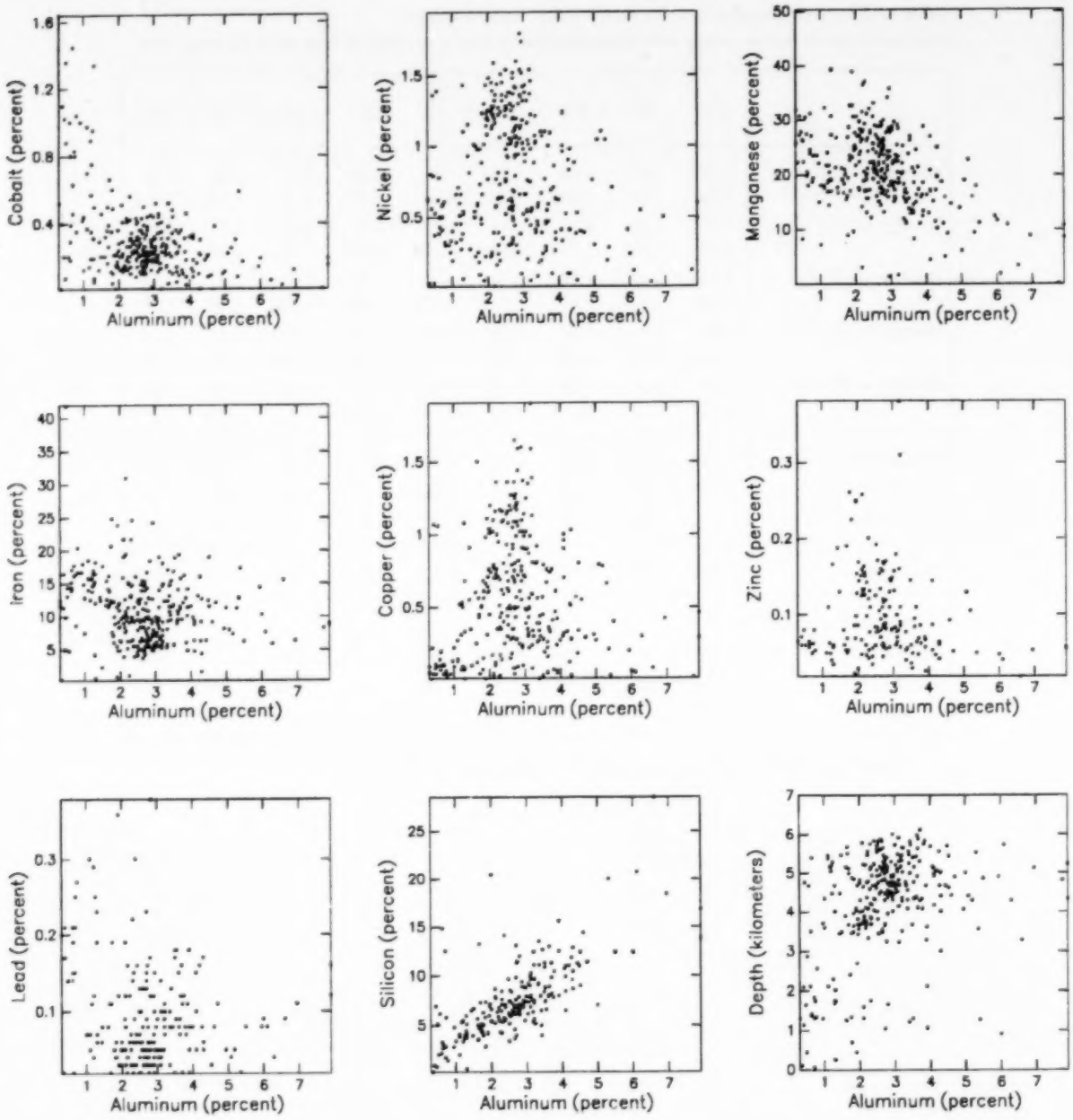

Figure 15. Scatter plots of Al content versus content of Co, Ni, Mn, Fe, Cu, Zn, Pb, and Si and depth of water above the sample site.

Table 8. Entropy computations for aluminum versus nine variables

[Parenthetical entries indicate entropy calculation normalized by 1/log *n*. de, depth of water above the sample site]

	Co	Ni	Mn	Fe	Cu	Zn	Pb	Si	de
Entropy of *X*	1.66 (.75)	1.67 (.76)	1.67 (.76)	1.67 (.76)	1.66 (.75)	1.59 (.72)	1.69 (.76)	1.65 (.75)	1.66 (.75)
Entropy of *Y*	1.47 (.61)	2.17 (.90)	1.78 (.74)	1.57 (.65)	2.05 (.85)	0.77 (.33)	1.67 (.72)	1.43 (.65)	1.95 (.81)
Entropy of (*X*,*Y*)	2.94 (.64)	3.67 (.80)	3.27 (.71)	3.10 (.67)	3.47 (.75)	2.26 (.50)	3.17 (.70)	2.66 (.60)	3.38 (.73)
Entropy of (*Y* | *X*)	1.28 (.53)	2.00 (.83)	1.60 (.67)	1.43 (.59)	1.80 (.75)	0.67 (.29)	1.48 (.64)	1.00 (.45)	1.72 (.71)
Entropy of (*X* | *Y*)	1.47 (.67)	1.49 (.68)	1.49 (.68)	1.53 (.69)	1.41 (.64)	1.49 (.67)	1.50 (.68)	1.22 (.55)	1.42 (.64)
Uncertainty (*Y* | *X*)	0.12	0.07	0.09	0.08	0.12	0.13	0.11	0.29	0.12
Uncertainty (*X* | *Y*)	.11	.10	.10	.08	.15	.06	.11	.25	.14
Uncertainty (*X*,*Y*)	.11	.09	.10	.08	.13	.08	.11	.27	.13
No. of points	301	310	311	310	305	178	203	211	303
Percent data content *									
Columns	100	100	100	100	100	88	88	88	100
Rows	90	100	81	72	90	50	100	77	100
Cells	37	56	48	42	47	20	42	35	55

* Percentage of rows, columns, and cells for which there are data.

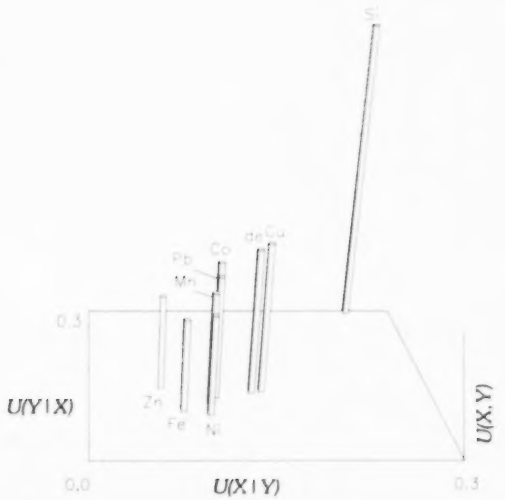

Figure 16. Perspective plot of uncertainties of Al versus uncertainties of Co, Ni, Mn, Fe, Cu, Zn, Pb, Si, and depth of water above the sample site (de). $U(Y \mid X)$, $U(X \mid Y)$, and $U(X,Y)$ are uncertainty coefficients.

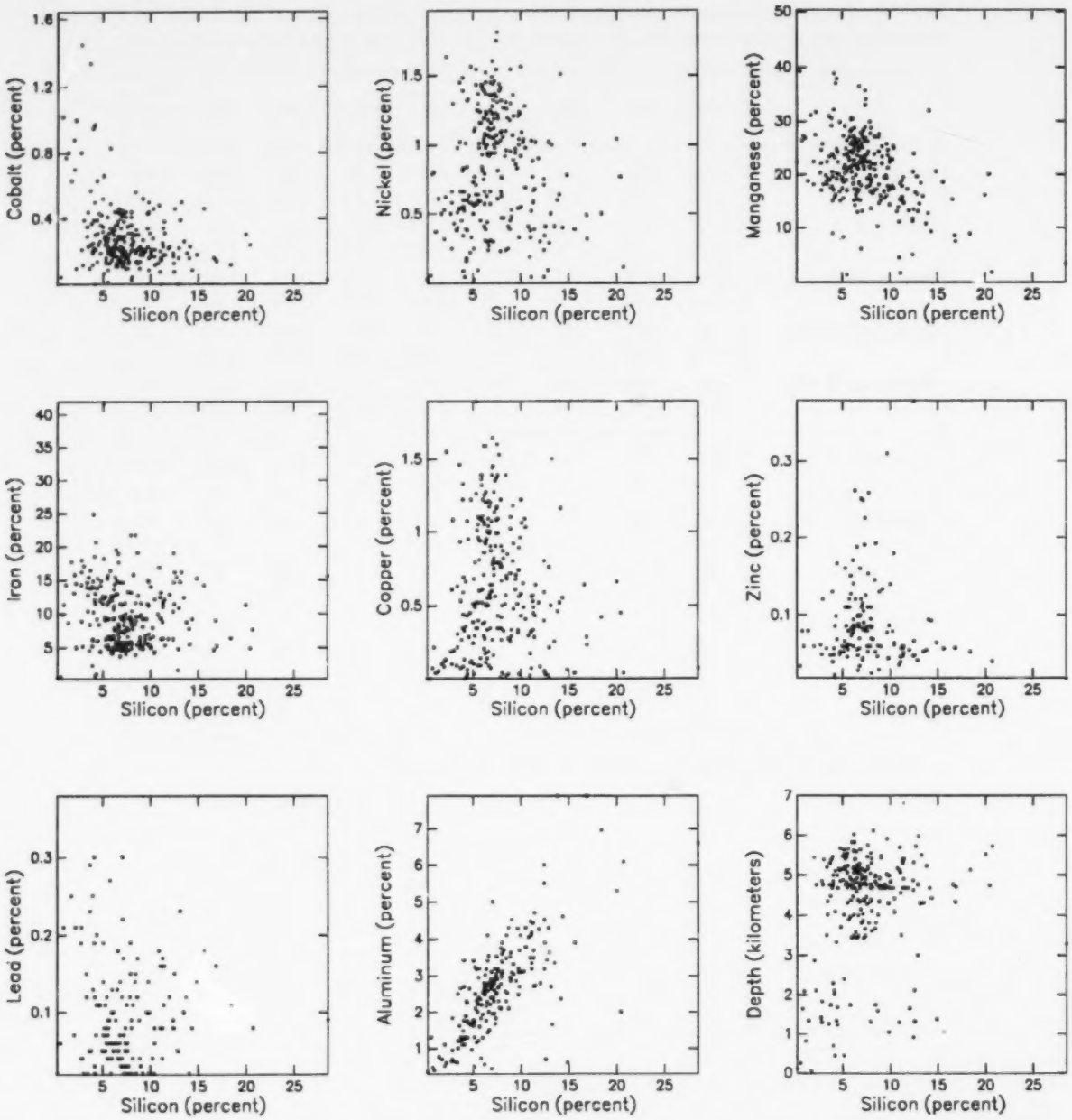

Figure 17. Scatter plots of Si content versus content of Co, Ni, Mn, Fe, Cu, Zn, Pb, and Al and depth of water above the sample site.

Table 9. Entropy computations for silicon versus nine variables

[Parenthetical entries indicate entropy calculation normalized by 1/log *n*. de, depth of water above the sample site]

	Co	Ni	Mn	Fe	Cu	Zn	Pb	Al	de
Entropy of *X*	1.44	1.45	1.45	1.45	1.45	1.43	1.51	1.43	1.45
	(.65)	(.66)	(.66)	(.66)	(.66)	(.65)	(.68)	(.65)	(.66)
Entropy of *Y*	1.38	2.18	1.71	1.44	2.15	0.63	1.71	1.65	1.81
	(.57)	(.91)	(.71)	(.60)	(.89)	(.27)	(.74)	(.75)	(.75)
Entropy of (*X*,*Y*)	2.69	3.46	2.98	2.79	3.45	2.02	3.07	2.66	3.09
	(.58)	(.75)	(.64)	(.60)	(.75)	(.45)	(.68)	(.60)	(.67)
Entropy of (*Y* | *X*)	1.24	2.00	1.52	1.33	1.99	0.58	1.56	1.22	1.64
	(.52)	(.83)	(.63)	(.55)	(.83)	(.25)	(.67)	(.55)	(.68)
Entropy of (*X* | *Y*)	1.31	1.27	1.26	1.34	1.29	1.39	1.35	1.00	1.28
	(.59)	(.58)	(.57)	(.61)	(.59)	(.63)	(.61)	(.45)	(.58)
Uncertainty (*Y* | *X*)	0.09	0.08	0.11	0.07	0.07	0.06	0.09	0.25	0.09
Uncertainty (*X* | *Y*)	.09	.12	.13	.07	.11	.03	.10	.29	.11
Uncertainty (*X*,*Y*)	.09	.09	.12	.07	.08	.04	.09	.27	.10
No. of points	258	266	267	266	261	144	156	211	254
Percent data content *									
Columns	77	77	77	77	77	77	77	77	77
Rows	81	100	81	63	90	40	90	88	100
Cells	30	48	35	31	46	16	41	35	45

* Percentage of rows, columns, and cells for which there are data.

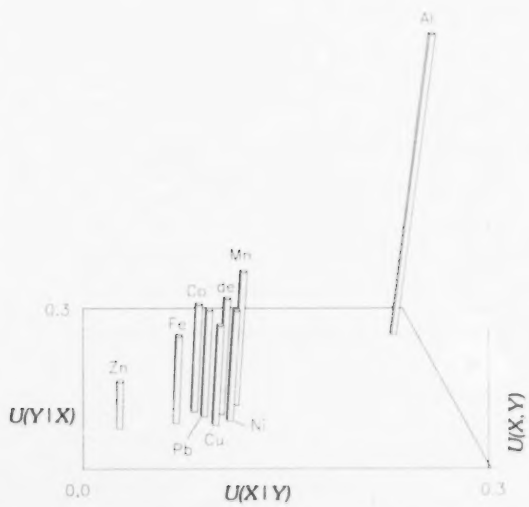

Figure 18. Perspective plot of uncertainties of Si versus uncertainties of Co, Ni, Mn, Fe, Cu, Zn, Pb, Al, and depth of water above the sample site (de). U(Y | X), *U*(*X* | *Y*), and *U*(*X*,*Y*) are uncertainty coefficients.

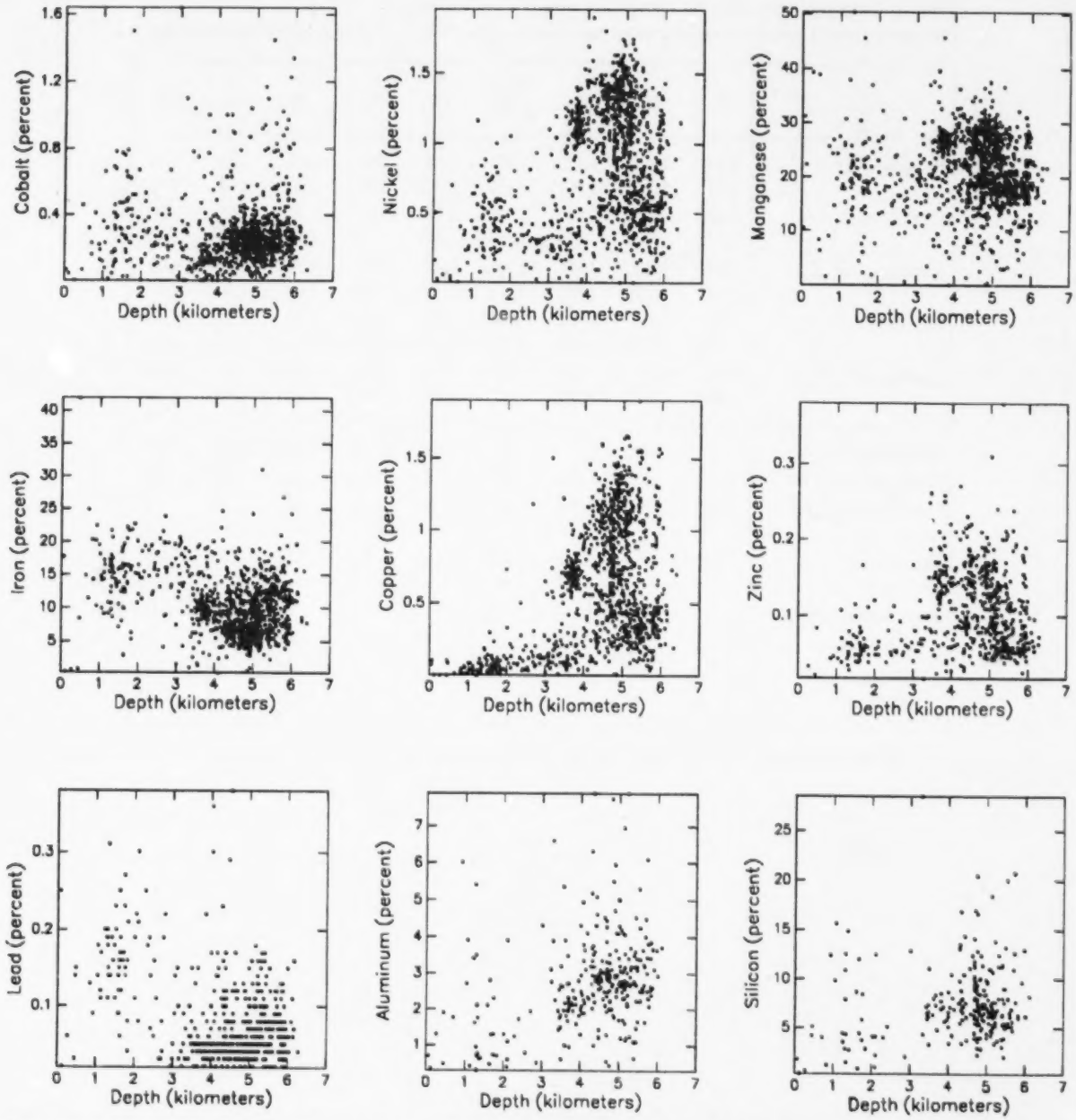

Figure 19. Scatter plots of depth of water above the sample site versus content of Co, Ni, Mn, Fe, Cu, Zn, Pb, Al, and Si.

Table 10. Entropy computations for depth versus nine variables

[Parenthetical entries indicate entropy calculation normalized by 1/log n. de, depth of water above the sample site]

	Co	Ni	Mn	Fe	Cu	Zn	Pb	Al	Si
Entropy of x	1.93	1.96	1.96	1.96	1.95	1.96	1.98	1.95	1.81
	(.80)	(.81)	(.81)	(.81)	(.81)	(.81)	(.82)	(.81)	(.75)
Entropy of y	1.35	2.18	1.76	1.50	2.15	0.76	1.46	1.66	1.45
	(.56)	(.91	(.73)	(.62)	(.89)	(.33)	(.63)	(.75)	(.66)
Entropy of (x,y)	3.16	3.89	3.59	3.24	3.72	2.60	3.22	3.38	3.09
	(.66)	(.81	(.74)	(.67)	(.77)	(.55)	(.68)	(.73)	(.67)
Entropy of $(y \mid x)$	1.23	1.93	1.63	1.27	1.76	0.64	1.24	1.42	1.28
	(.51)	(.80	(.68)	(.53)	(.73)	(.27)	(.54)	(.64)	(.58)
Entropy of $(x \mid y)$	1.81	1.71	1.83	1.73	1.56	1.84	1.76	1.72	1.64
	(.75)	(.71	(.76)	(.72)	(.65)	(.76)	(.73)	(.71)	(.68)
Uncertainty $(y \mid x)$	0.08	0.11	0.07	0.15	0.17	0.15	0.14	0.14	0.11
Uncertainty $(x \mid y)$	.06	.12	.06	.11	.19	.06	.10	.12	.09
Uncertainty (x,y)	.07	.12	.06	.13	.18	.08	.12	.13	.10
No of points	1,058	1,118	1,108	1,100	1,107	646	616	303	254
Percent data content *									
Columns	100	100	100	100	100	100	100	100	100
Rows	100	100	90	72	90	50	100	100	77
Cells	55	67	69	53	52	23	58	55	45

* Percentage of rows, columns, and cells for which there are data.

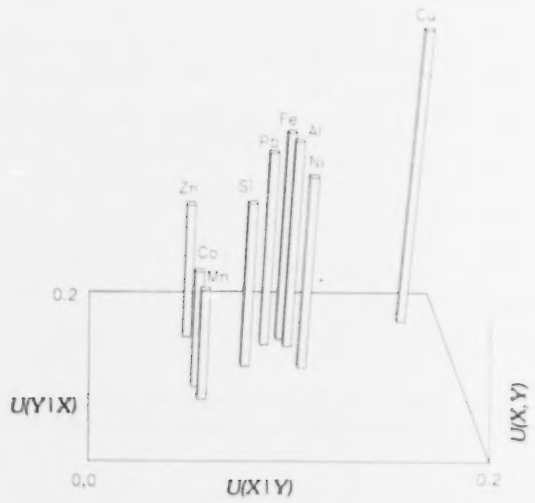

Figure 20. Perspective plot of uncertainties of depth of water above the sample site (de) versus uncertainties of Co, Ni, Mn, Fe, Cu, Zn, Pb, Al, and Si. $U(Y \mid X)$, $U(X \mid Y)$, and $U(X,Y)$ are uncertainty coefficients.

Table 11. Correlation coefficients of raw data

[Cophenetic correlation coefficient = 0.87. See p. 5 for discussion]

	Co	Ni	Mn	Fe	Cu	Zn	Pb	Al	Si	Depth
Co	1.00	-0.18	0.11	0.38	-0.33	-0.11	0.60	-0.42	-0.31	-0.40
Ni	-.18	1.00	.71	-.66	.89	.04	-.42	-.07	-.09	.27
Mn	.11	.71	1.00	-.41	.60	.11	-.12	-.45	-.45	-.02
Fe	.38	-.66	-.41	1.00	-.70	-.18	.54	-.21	-.07	-.34
Cu	-.33	.89	.60	-.70	1.00	.05	-.53	.04	-.06	.42
Zn	-.11	.04	.11	-.18	.05	1.00	-.13	-.15	-.17	-.11
Pb	.60	-.42	-.12	.54	-.53	-.13	1.00	-.19	-.11	-.44
Al	-.42	-.07	-.45	-.21	.04	-.15	-.19	1.00	.70	.33
Si	-.31	-.09	-.45	-.07	-.06	-.17	-.11	.70	1.00	.18
Depth	-.40	.27	-.02	-.34	.42	-.11	-.44	.33	.18	1.00

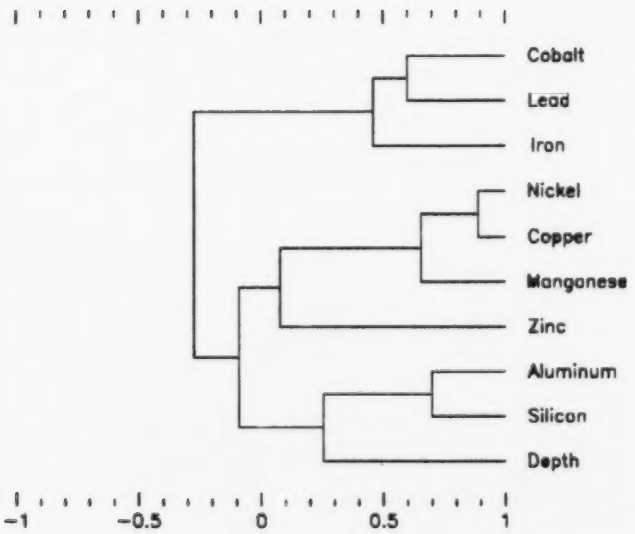

Figure 21. Dendritic diagram showing clusters of variables based on linear correlation coefficients of raw data.

Table 12. Correlation coefficients of logarithms of raw data

[Cophenetic correlation coefficient = 0.83. See p. 5 for discussion]

	Co	Ni	Mn	Fe	Cu	Zn	Pb	Al	Si	Depth
Co	10.00	0.18	0.31	0.37	-0.11	-0.17	0.51	-0.36	-0.15	-0.12
Ni	.18	1.00	.70	-.42	.80	.59	-.29	.05	.11	.36
Mn	.31	.70	1.00	-.22	.45	.50	-.07	-.32	-.38	.05
Fe	.37	-.42	-.22	1.00	-.48	-.51	.51	-.12	.09	-.19
Cu	-.11	.80	.45	-.48	1.00	.54	-.48	.28	.22	.63
Zn	-.17	.59	.50	-.51	.54	1.00	-.51	-.14	-.28	.10
Pb	.51	-.29	-.07	.51	-.48	-.51	1.00	-.19	-.06	-.32
Al	-.36	.05	-.32	-.12	.28	-.14	-.19	1.00	.71	.46
Si	-.15	.11	-.38	.09	.22	-.28	-.06	.71	1.00	.44
Depth	-.12	.36	.05	-.19	.63	.10	-.32	.46	.44	1.00

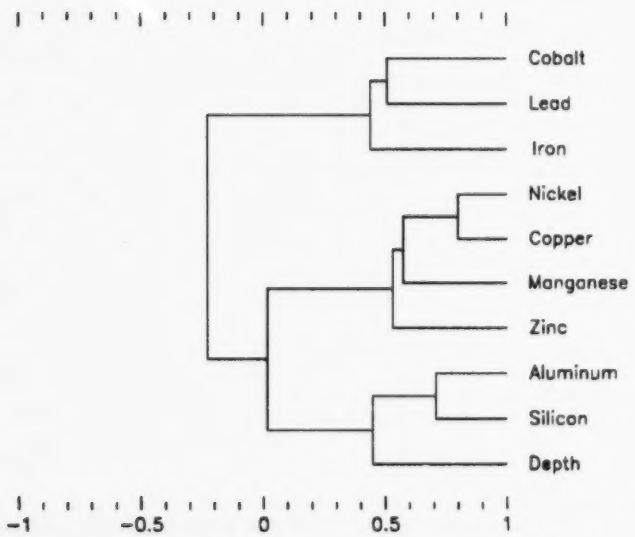

Figure 22. Dendritic diagram showing clusters of variables based on linear correlation coefficients of logarithms of the raw data.

Table 13. Symmetrical uncertainties of elemental enrichments

[Cophenetic correlation coefficient = 0.86. See p. 5 for discussion]

	Co	Ni	Mn	Fe	Cu	Zn	Pb	Al	Si	Depth
Co	1.00	0.06	0.09	0.09	0.09	0.04	0.15	0.09	0.07	0.03
Ni	.06	1.00	.18	.12	.27	.12	.08	.07	.08	.05
Mn	.09	.18	1.00	.09	.14	.11	.05	.12	.12	.02
Fe	.09	.12	.09	1.00	.13	.11	.17	.06	.06	.03
Cu	.09	.27	.14	.13	1.00	.10	.11	.10	.08	.04
Zn	.04	.12	.11	.11	.10	1.00	.09	.06	.04	.02
Pb	.15	.08	.05	.17	.11	.09	1.00	.08	.09	.03
Al	.09	.07	.12	.06	.10	.06	.08	1.00	.23	.09
Si	.07	.08	.12	.06	.08	.04	.09	.23	1.00	.08
Depth	.03	.05	.02	.03	.04	.02	.03	.09	.08	1.00

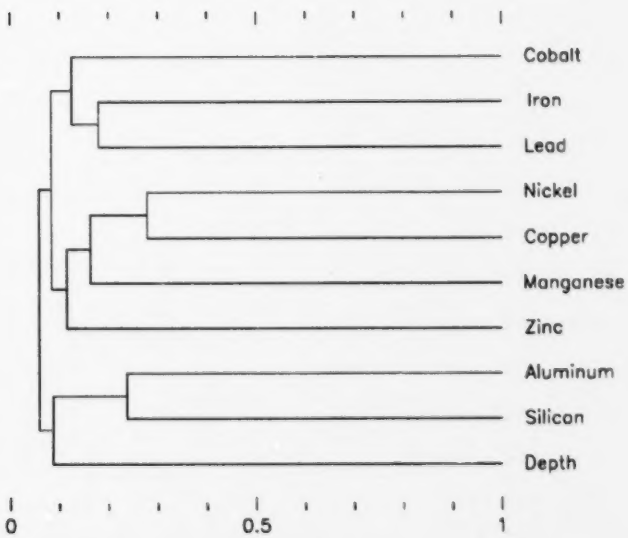

Figure 23. Dendritic diagram showing clusters of variables based on uncertainty coefficients of elemental enrichments.

Table 14. Symmetrical uncertainties of raw data

[Cophenetic correlation coefficient = 0.88. See p. 5 for discussion]

	Co	Ni	Mn	Fe	Cu	Zn	Pb	Al	Si	Depth
Co	1.00	0.08	0.08	0.10	0.10	0.06	0.17	0.11	0.09	0.07
Ni	.08	1.00	.26	.23	.36	.22	.12	.09	.09	.12
Mn	.08	.26	1.00	.13	.19	.23	.09	.10	.12	.06
Fe	.10	.23	.13	1.00	.26	.15	.18	.08	.07	.13
Cu	.10	.36	.19	.26	1.00	.17	.16	.13	.08	.18
Zn	.06	.22	.23	.15	.17	1.00	.17	.08	.04	.08
Pb	.17	.12	.09	.18	.16	.17	1.00	.11	.09	.12
Al	.11	.09	.10	.08	.13	.08	.11	1.00	.27	.13
Si	.09	.09	.12	.07	.08	.04	.09	.27	1.00	.10
Depth	.07	.12	.06	.13	.18	.08	.12	.13	.10	1.00

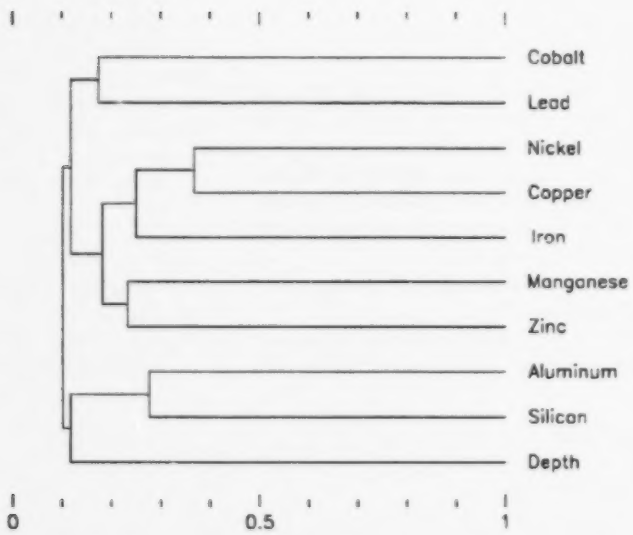

Figure 24. Dendritic diagram showing clusters of variables based on uncertainty coefficients of the raw data.

Table 15. Cluster analysis of linear correlation coefficients of raw data and their logarithmic transforms and uncertainty coefficients of elemental enrichments and raw data

Cluster	Correlation coefficient		Uncertainty coefficient	
	Raw data	Log raw data	Enrichment	Raw data
I	Cobalt Lead Iron	Cobalt Lead Iron	Iron Lead Cobalt	Cobalt Lead
II	Nickel Copper Manganese Zinc	Nickel Copper Manganese Zinc	Nickel Copper Manganese Zinc	Nickel Copper Iron Manganese Zinc
III	Aluminum Silicon Depth	Aluminum Silicon Depth	Aluminum Silicon Depth	Aluminum Silicon Depth

REFERENCES CITED

Botbol, J.M., 1987, Geochemical enrichment vectors applied to manganese oxide phases in the northern Pacific: U.S. Geological Survey Open-File Report 87–197, 17 p.

Botbol, J.M., and Evenden, G.I., in press, Descriptive statistics and spatial distributions of geochemical variables associated with manganese oxide-rich phases in the northern Pacific: U.S. Geological Survey Bulletin 1863.

Davis, J.C., 1973, Statistics and data analysis in geology: New York, John Wiley and Sons, 550 p.

Frazer, J.Z., and Fisk, M.B., 1980, Availability of copper, nickel, cobalt, and manganese from ocean ferromanganese nodules, pt. III: La Jolla, Calif., Scripps Institution of Oceanography, SIO reference 80–16 [prepared for U.S. Bureau of Mines], 117 p.

Huntsberger, D.V., 1963, Elements of statistical inference: Boston, Allyn and Bacon, 291 p.

McKelvey, V.E., Wright, N.A., and Bowen, R.W., 1983, Analysis of the world distribution of metal-rich subsea manganese nodules: U.S. Geological Survey Circular 886, 55 p.

Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T., 1988, Numerical recipes in C—The art of scientific computing: Cambridge, England, Cambridge University Press, p. 498.

Singh, A., 1966, Great ideas in information theory, language, and cybernetics: New York, Dover Publications, 338 p.

SELECTED SERIES OF U.S. GEOLOGICAL SURVEY PUBLICATIONS

Periodicals

Earthquakes & Volcanoes (issued bimonthly).
Preliminary Determination of Epicenters (issued monthly).

Technical Books and Reports

Professional Papers are mainly comprehensive scientific reports of wide and lasting interest and importance to professional scientists and engineers. Included are reports on the results of resource studies and of topographic, hydrologic, and geologic investigations. They also include collections of related papers addressing different aspects of a single scientific topic.

Bulletins contain significant data and interpretations that are of lasting scientific interest but are generally more limited in scope or geographic coverage than Professional Papers. They include the results of resource studies and of geologic and topographic investigations; as well as collections of short papers related to a specific topic.

Water-Supply Papers are comprehensive reports that present significant interpretive results of hydrologic investigations of wide interest to professional geologists, hydrologists, and engineers. The series covers investigations in all phases of hydrology, including hydrogeology, availability of water, quality of water, and use of water.

Circulars present administrative information or important scientific information of wide popular interest in a format designed for distribution at no cost to the public. Information is usually of short-term interest.

Water-Resources Investigations Reports are papers of an interpretive nature made available to the public outside the formal USGS publications series. Copies are reproduced on request unlike formal USGS publications, and they are also available for public inspection at depositories indicated in USGS catalogs.

Open-File Reports include unpublished manuscript reports, maps, and other material that are made available for public consultation at depositories. They are a nonpermanent form of publication that may be cited in other publications as sources of information.

Maps

Geologic Quadrangle Maps are multicolor geologic maps on topographic bases in 7 1/2- or 15-minute quadrangle formats (scales mainly 1:24,000 or 1:62,500) showing bedrock, surficial, or engineering geology. Maps generally include brief texts; some maps include structure and columnar sections only.

Geophysical Investigations Maps are on topographic or planimetric bases at various scales; they show results of surveys using geophysical techniques, such as gravity, magnetic, seismic, or radioactivity, which reflect subsurface structures that are of economic or geologic significance. Many maps include correlations with the geology.

Miscellaneous Investigations Series Maps are on planimetric or topographic bases of regular and irregular areas at various scales; they present a wide variety of format and subject matter. The series also includes 7 1/2-minute quadrangle photogeologic maps on planimetric bases which show geology as interpreted from aerial photographs. Series also includes maps of Mars and the Moon.

Coal Investigations Maps are geologic maps on topographic or planimetric bases at various scales showing bedrock or surficial geology, stratigraphy, and structural relations in certain coal-resource areas.

Oil and Gas Investigations Charts show stratigraphic information for certain oil and gas fields and other areas having petroleum potential.

Miscellaneous Field Studies Maps are multicolor or black-and-white maps on topographic or planimetric bases on quadrangle or irregular areas at various scales. Pre-1971 maps show bedrock geology in relation to specific mining or mineral-deposit problems; post-1971 maps are primarily black-and-white maps on various subjects such as environmental studies or wilderness mineral investigations.

Hydrologic Investigations Atlases are multicolored or black-and-white maps on topographic or planimetric bases presenting a wide range of geohydrologic data of both regular and irregular areas; principal scale is 1:24,000 and regional studies are at 1:250,000 scale or smaller.

Catalogs

Permanent catalogs, as well as some others, giving comprehensive listings of U.S. Geological Survey publications are available under the conditions indicated below from the U.S. Geological Survey, Books and Open-File Reports Section, Federal Center, Box 25425, Denver, CO 80225. (See latest Price and Availability List.)

"Publications of the Geological Survey, 1879-1961" may be purchased by mail and over the counter in paperback book form and as a set of microfiche.

"Publications of the Geological Survey, 1962-1970" may be purchased by mail and over the counter in paperback book form and as a set of microfiche.

"Publications of the U.S. Geological Survey, 1971-1981" may be purchased by mail and over the counter in paperback book form (two volumes, publications listing and index) and as a set of microfiche.

Supplements for 1982, 1983, 1984, 1985, 1986, and for subsequent years since the last permanent catalog may be purchased by mail and over the counter in paperback book form.

State catalogs, "List of U.S. Geological Survey Geologic and Water-Supply Reports and Maps For (State)," may be purchased by mail and over the counter in paperback booklet form only.

"Price and Availability List of U.S. Geological Survey Publications," issued annually, is available free of charge in paperback booklet form only.

Selected copies of a monthly catalog "New Publications of the U.S. Geological Survey" available free of charge by mail or may be obtained over the counter in paperback booklet form only. Those wishing a free subscription to the monthly catalog "New Publications of the U.S. Geological Survey" should write to the U.S. Geological Survey, 582 National Center, Reston, VA 22092.

Note.--Prices of Government publications listed in older catalogs, announcements, and publications may be incorrect. Therefore, the prices charged may differ from the prices in catalogs, announcements, and publications.